JMIR mHealth and uHealth

Impact Factor (2024): 5.4
Volume 10 (2022), Issue 4 ISSN 2291-5222 Editor in Chief: Lorraine Buis, PhD, MSI

Contents

Viewpoints

Reviews

Original Papers

Viewpoint

Development of a Mobile App for Clinical Research: Challenges and Implications for Investigators

Shibani Chettri[1], MPH; Vivian Wang[2], BSPH; Eli Asher Balkin[3], MS; Michael F Rayo[3], PhD; Clara N Lee[4,5,6], MD, MPP

[1]College of Public Health, The Ohio State University, Columbus, OH, United States

[2]College of Medicine, Central Michigan University, Mount Pleasant, MI, United States

[3]Department of Integrated Systems Engineering, The Ohio State University, Columbus, OH, United States

[4]Department of Plastic and Reconstructive Surgery, College of Medicine, The Ohio State University, Columbus, OH, United States

[5]Division of Health Services Management and Policy, College of Public Health, The Ohio State University, Columbus, OH, United States

[6]The James Comprehensive Cancer Center-Solove Research Institute, The Ohio State University, Columbus, OH, United States

Corresponding Author:
Shibani Chettri, MPH
College of Public Health
The Ohio State University
1529 N High St, Apt 614
Columbus, OH, 43201
United States
Phone: 1 2404297749
Email: Chettri.1@osu.edu

Abstract

Advances in mobile app technologies offer opportunities for researchers to feasibly collect a large amount of patient data that were previously inaccessible through traditional clinical research methods. Collection of data via mobile devices allows for several advantages, such as the ability to continuously gather data outside of research facilities and produce a greater quantity of data, making these data much more valuable to researchers. Health services research is increasingly incorporating mobile health (mHealth), but collecting these data in current research institutions is not without its challenges. Our paper uses a specific example to depict specific challenges of mHealth research and provides recommendations for investigators looking to incorporate digital app technologies and patient-collected digital data into their studies. Our experience describes how clinical researchers should be prepared to work with variable software and mobile app development timelines; research institutions that are interested in participating in mHealth research need to invest in supporting information technology infrastructures in order to be a part of the growing field of mHealth and gain access to valuable patient-collected data.

(JMIR Mhealth Uhealth 2022;10(4):e32244) doi:10.2196/32244

KEYWORDS

mHealth; mobile app; patient-collected data; data security; mobile health; patient data; clinical research; research facilities

Introduction

The rapid adoption of internet-connected digital devices is facilitating a change in how health care providers can monitor patient health. The proliferation of devices, such as cellular phones, tablet computers, and smartwatches, permits clinicians to gather data in ways that were previously impossible. Not only do these devices produce more data, but they also produce more useful data since they can be collected while patients are attending to their daily lives.

As more mobile technology and devices become available, health services research is also evolving [1,2]. Research on the delivery of health care no longer needs to restrict data collection to health care facilities or to the timing of clinical schedules. This proliferation of new research tools has also dramatically changed the roles and skills required of modern clinical and health services researchers. Such investigators must become accustomed to research with a substantial medical informatics component.

This paper describes the challenges that mobile health (mHealth) research can pose for clinical researchers, including unique issues with investigator roles, patients as researchers, data

security, contracting with app developers, and informatics infrastructure. We use the example of a specific research study to illustrate these challenges and provide recommendations.

The Rise of mHealth

Three technological trends have facilitated the shift toward mHealth: daily carry, reliable connectivity, and near-universal usage [3,4]. The consistent use of cell phones, tablets, and smartwatches by people throughout the day for web browsing, email, music, calls, and navigation means that these devices are in close proximity to potential research participants for most of the day. Thus, health care research can be conducted using data collected by these devices, no longer requiring patients to carry additional devices or modify their daily routines [5-8].

Predictable internet connectivity allows devices to automatically send patient data to health care providers and researchers without the need for patient intervention. This reduces the burden of participation for the patient, while permitting providers to review the data in preparation for the patient's next scheduled appointment. Medical staff can also monitor the patient from afar and contact the patient for intervention if necessary.

The nearly universal use of internet-connected devices, especially cell phones, makes population-based data collection more feasible, which has important implications for health equity. Specifically, populations that have been left out of prior research may become more accessible. Finally, it also opens up possibilities for dissemination and implementation of interventions developed using mHealth, thus facilitating a wider experience of benefits at a lowered preparatory and operational cost.

WORDS: An Illustration

The Women and Oncologists Reaching Decisions about Surgery (WORDS) study sought to understand how women with breast cancer make decisions about surgery, and how they communicate with their clinicians about treatment options. Study participants recorded their conversations with surgical oncology, radiation oncology, and plastic surgery clinicians using RecorDr, a smartphone app we created. Because of the multiple patient-provider conversations across multiple locations that is typical in determining the optimal care plans for these patients [9], our patient-as-researcher method would be particularly valuable.

The primary outcomes of the WORDS study were (1) to develop and test a mobile app in a clinical setting for research and (2) to evaluate communication and decision-making about contralateral prophylactic mastectomy. The WORDS study had a total of 105 participants. A control group was not used.

The benefit of using our mobile app was that data collection would not need a research assistant at each appointment. In effect, the patients were deputized as research assistants, responsible for recording their encounters with physicians. They went through a brief orientation during their first visit to the cancer center, but then they could make additional recordings at other visits, which facilitated data collection at dates, times, and locations where the formal research team could not participate. The app also enabled patients to record conversations with family members if they chose to do so. Such broad data collection overcomes the prior obstacles of many communication studies, provides a more complete picture of the decision-making process, and increases the generalizability of research to nonacademic settings.

In order to deputize these patients and realize these benefits, our study required a substantial informatics component. However, most clinical and health services researchers do not have informatics expertise [10]. This highlights the importance of team science and collaborating with investigators who provide that expertise. Even then, the principal investigator (PI) must be familiar enough with the particular area to communicate effectively across the assembled disciplines, including software design, software development, data infrastructure, and data security.

Within our study there were three main teams: the clinical research team, the systems engineering team, and the information technology (IT) team. The PI was a decision scientist, health services researcher, and surgeon, with expertise in decision-making about breast cancer treatments. The clinical team provided the clinical context for the app design and oversaw the patient and clinician recruitment procedures. The engineering team oversaw the app design, app development, and data transfer procedures. The clinical team relied on the engineering team and the developer. The clinical and engineering teams entrusted the medical center's IT department to develop the infrastructure and procedures for secure data transfer and storage.

Investigator Roles

As investigators begin to use new technological developments, they face a foundational choice between acquiring the new skills themselves or collaborating with a larger research team comprising experts in the required disciplines. In mHealth, the former option of learning about computer programming, mobile app development, and digital data storage is not highly feasible for most clinical researchers. The latter option, the more common choice and the hallmark of modern team-based science, requires research collaborators with the required expertise and willingness to collaborate. In the case of mHealth programs, collaborators often hail from such fields as engineering, computer science, and informatics. Clinical investigators doing this type of research, however, should acquire enough basic knowledge about app development and informatics to collaborate meaningfully with these parties.

We recommend that academic centers engaging in mHealth research have "common-pool" resources available for all researchers to use, such as in-house methodological consulting to aid in the development of practices and procedures for the use of mHealth technologies. The mHealth shared resource at the Stephenson Cancer Center at the University of Oklahoma is an example of a common-pool site that houses a variety of important mHealth technologies as well as connections to clinical informatics experts [11]. Additionally, short courses

for researchers on contracting and working with programmers would aid in the process.

Institutions with infrastructure to support mHealth research will reduce the amount of time spent by research teams figuring out who in the organization is properly positioned to offer the services needed, or at least to inform the team that those services could not be offered. Providing clarity to researchers regarding which support services are currently available and how to access them would increase the development and adoption of mHealth research tools. This would allow research teams to know early on whether a project is not feasible within the existing infrastructure, thus saving them time and resources. In our experience, although our research project was approved to take place at our institution, we only found out several months later that some of the necessary components of our study, such as a secure data storage system and developers who had familiarity with mobile app data security, were not readily available to our research team through our university. We then spent a significant amount of time looking into third parties who provided these services. We suggest that researchers familiarize themselves with the mHealth resources at their institutions and consider the availability of an internal mobile app developer who is familiar with mHealth research, a dedicated IT department that is familiar with the needs of mHealth researchers, and the mobile data security measures currently in place at the institution. A successful model offers these services at low or no cost to research teams. Universities that choose to invest in integrated clinical informatics centers will be better equipped to keep up with the fast-paced changes in mHealth. Research and researchers who inquire about these services will be better able to anticipate the logistics, hurdles, and feasibility of their mHealth projects.

Participants as Deputized Researchers: New Challenges

The involvement of research participants as data collectors creates new challenges. These new initiates will generally have far less training than a typical research assistant, will not be under the employ of the research team, and will generally need some short- or long-term benefit for participating [12]. Usability issues regarding study tools that a typical research team could easily address can completely derail these new, decentralized teams. We conducted two rounds of informal, formative usability testing before our first app release, and we had to quickly revise the app after the first release after new usability issues were revealed. The team had to be constantly vigilant, as the participants would not necessarily signal when they could not or would not complete their tasks; instead, they often dropped out. We provided financial reimbursement to participants, but only at the point of enrollment rather than after each data collection point. We also provided a nonfinancial incentive to mitigate participant dropout, allowing participants to keep their recordings and including a bookmark feature. This feature enabled them to tap their screen whenever something occurred that they would want to refer to later. In this way, their work as a researcher was serving their needs as a patient.

We suggest that when conducting research that includes participants as data collectors, researchers should minimize the burden for participants. If feasible, researchers should schedule financial reimbursements after the completion of different stages of data collection (eg, after successful recording). Apps should be designed to provide a nonfinancial benefit to aid in participation and retention. Additionally, usability testing in focus groups that closely resemble the demographics of the target population of the study will reduce the need for large revisions after the release of the mobile app. Although comfort with mobile technology may vary across demographics, participants who use mobile internet data technologies will likely also be comfortable with using mobile technologies in mHealth studies [13]. Clinical researchers should also consider simulating the characteristics of the setting in which the app will be used and take into account factors such as time constraints of clinical schedules, participants being in an unfamiliar setting, and needing to connect to a different Wi-Fi network than usual. Teams should also be prepared to conduct usability testing in the actual study setting that is as similar as possible to the actual setting, such as in clinic and with actual patients who are receiving care rather than just with cancer survivors who have completed treatment or healthy volunteers. Additionally, researchers should touch base with participants throughout the project to get feedback about which technological processes are working and which are problematic, so that problems can be addressed in real time. We tracked participant uploading of recordings to the server, and we suggest such monitoring after the launch of the app in order to incorporate an automatic notification system to assess for usability issues. Such feedback could be obtained via targeted inquiries in the app itself, using metrics such as number of data files received, and less time spent on the app. The app itself could then target users who may be using the app less frequently or are encountering usability problems. Other feedback options include surveys, phone calls, or in-person inquiries by on-site research personnel. However, such approaches add burden to study participants, when a major goal of mHealth research is to minimize participant burden and research staff involvement.

Data Security in mHealth

Overview

As mHealth grows, researchers will have to navigate data security concerns for more types of data and a wider range of situations. Since there are few industry-wide standards, each project and institution may need to develop its own approach.

The Health Insurance Portability and Accountability Act

Questions are arising about what constitutes protected health information (PHI) under the Health Insurance Portability and Accountability Act. Initially, the Institutional Review Board (IRB) approved our study as not involving PHI. They considered the voice recordings not identifiable. Thus, we considered commercially available data storage options, such as Amazon Web Services (AWS) and Dropbox, as well as university-managed options, such as Enterprise Box and the internally owned and operated medical center storage server.

After taking into consideration cost, reliability, and security, we selected Enterprise Box, which was already implemented in our university and medical center. However, while the project was underway, the university and medical center security offices changed their interpretation of the institutional data classification standards, deeming voice data as PHI.

This change, coupled with security concerns of nonuniversity devices (ie, the mobile devices of our patients), meant that a wholly new data storage solution had to be designed and implemented. The solution consisted of a new secure medical center server for storage and secure file transfer protocol (SFTP) for data transfer from the app to the server. Setting up the server and SFTP involved a partnership between our team, the medical center IT department, and the university and medical center security office. Each of these steps, including evaluating the problem (ie, audio recordings as PHI), identifying solutions (ie, server and SFTP), and implementing the solution (ie, setup of the server and SFTP) took several months, with the final implementation completed 1 year after the storage problem was identified.

Ultimately, three important data management lessons emerged regarding data transfer, data storage, and participant data security.

Data Transfer

Data collected on a mobile app need to be vetted before being transferred to an institutional server. Additionally, data collected on personal patient devices need to be screened, since the recordings are made outside the jurisdiction of the medical center. We sent all files from mobile devices by SFTP to a server environment controlled by our institution but outside of the firewall (ie, a DMZ [demilitarized zone] network), removing the ability of malicious files or actors to infiltrate the secured institutional network. Only specific processes on specific ports within the secured network were allowed access to the file system in the DMZ network. These processes periodically moved the patient conversation audio files between the DMZ network and the secured network.

Data Storage

When selecting a modality for PHI storage, researchers need to consider the security level of their storage system. Commercially available storage options, such as AWS, Dropbox, and Box, may not meet the institutional or regulatory security standards to house PHI; thus, other options need to be used. As new forms of patient data continue to emerge, institutional classifications of these new data may not yet be determined at academic centers. We recommend that teams encountering similar uncertainty regarding patient data classification handle patient data as PHI to prevent the need to change security standards after projects are active.

Participant Data Security

By collecting data on portable, user-controlled devices, the data are only as secure as the device itself. Recording data are vulnerable since patients have all their recordings on their smartphones. If a patient loses their phone, recordings could be accessed by others. Researchers should be prepared to train patients in security hygiene, such as making sure mobile devices are password protected and being aware of where they leave their devices. Additional considerations to minimize privacy and security threats include training participants to be familiar with the type of data being collected and the use of the collected data. Patients should also be reminded to prohibit data access over unsecure Wi-Fi networks or hot spots [14]. Although researchers can take these precautions, a data breach on the patient's personal device remains a risk, over which investigators have limited control.

Contracting With an Independent App Developer

Much of mHealth research requires coordination between the core research team and an app developer. Most app developers are accustomed to creating apps on a contractual basis for monetary compensation and not necessarily for research. They may be unfamiliar with the flow and pace of clinical research and the importance of flexibility, especially when working with a protocol office, IRB, or medical center information security team. App development for clinical research requires stringent data security and privacy provisions to ensure that patient data are protected. We identified a professional developer through our professional network and negotiated a contract based on what we thought would be needed. However, requirements of the medical center IT department and the IRB required multiple changes and increasing complexity to the app. We found that our first developer was not accustomed to this level of complexity and could not accommodate multiple changes. We changed developers, which resulted in delays associated with hiring and onboarding a new developer. Our second developer tried to build off of the existing app for several months but eventually concluded that building from scratch would be more effective. The second developer was a computer science student who had the necessary skills and flexibility for the project.

We recommend that institutions have their own app developers with mHealth research experience, and that they create infrastructure to support hiring of external app developers when needed. The mHealth center at the University of Pennsylvania offers such support by searching for and preapproving app developers on behalf of research teams qualified to take on clinical projects with a strong informatics component. This saves the research team valuable time and resources [15]. When selecting a mobile app developer, researchers should seek those with experience related to research and implementation, as opposed to creating proofs of concept or demos only. For future studies, our goal is to design a team that is composed of specialists in app development, user interface and user experience design, and security. While it would be preferable for the developer to have, or be trained to have, sufficient understanding of security and other relevant areas, this may be difficult to find. Instead, we suggest looking for developers with experience building iOS and Android natively, as opposed to using libraries' cross-platform software. They will be more likely to be able to create apps that are fit to transfer large data files quickly and securely. Additionally, it is crucial to make clear the scale of the app and the expectation for the end product

app. We also recommend that investigators consider the levels of expertise, skill, and flexibility over the level of formal education.

Conducting Research Within an Evolving IT Infrastructure

mHealth research may require working within an IT infrastructure that was not designed to support research. At many medical centers, IT staff members are more familiar with the clinical operational IT needs of the medical center, rather than with informatics research. Medical center IT departments may assign research projects lower priority than urgent, daily clinical IT needs. For our study, our institution did not have a standardized protocol for informatics research at the time. We were not assigned one person from the IT department to be dedicated to our project, leaving no clear point of contact or consistent approach to making progress. For example, when we sought to develop a firewalled storage solution for our data, we had challenges in clearly communicating this need and determining which types of expertise were required. We depended on our IT collaborators to identify the right types of expertise required. However, they were not necessarily familiar with the urgency of the project, which was related to the award period of the study's funding; the usability of the mobile app in the clinic; or security issues for audio PHI from patient mobile devices.

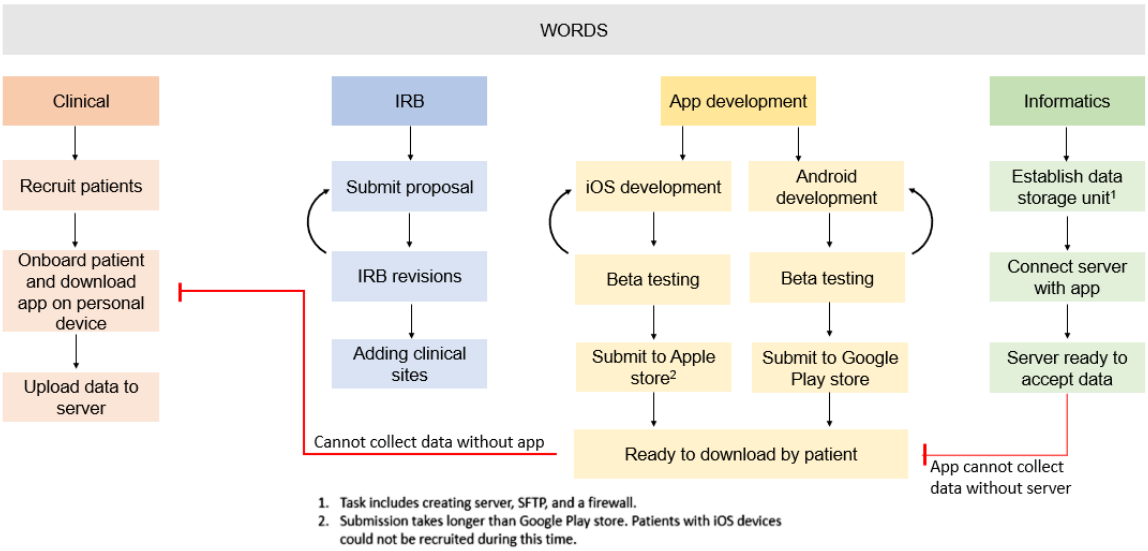
Currently, many universities do not have the resources and infrastructure necessary to keep up with the demand for informatics integration into clinical research [16,17]. We contacted multiple institutions and found only a few that had mHealth-related development cores. Most institutions reported that they rely on medical center IT and IT security to help with server-side issues and contract out software development to third-party companies. However, informatics plays an integral part in advancing modern research, and its contribution will only grow in the future. Retail and service-oriented industries are using smartphones and other technology to tailor their products and accessibility to their customers. Similarly, technology could be leveraged to not only gather data from patients but empower patients to be data collectors in their own right, thereby potentially enhancing their experience in health care settings while also improving the quality of the data received.

We recommend that institutions develop informatics infrastructure specific to mHealth research. This would include technologies such as on-site mHealth experts and support personnel, standardized mHealth procedures, and improved data security infrastructure. A designated department should be established for mHealth and IT-dependent research that is distinct from IT for the clinical enterprise. A designated department will streamline projects by housing all mobile app research experts in one place and benefit from lessons learned across projects. Our institution now has a research IT department, which is developing infrastructure and standardized procedures. The University of North Carolina at Chapel Hill has the Connected Health for Applications and Interventions Core, which serves as a centralized support center for investigators seeking to incorporate mobile technology, user inquiry, and graphic design into their research [18]. Similarly, Duke University's Mobile App Gateway is a "one-stop shop" for assistance in the development of research-focused mobile apps [19]. It offers assistance with locating and contacting the appropriate IT staff, support with external vendors, contracting, and preapproved app-related language for IRB submissions [19]. It offers introductory design training workshops for researchers interested in using mHealth technologies, but who do not have a background in digital design.

Interactions in a Complex System

As with many large research projects, attempting to explore the larger project process through deconstructive processes (ie, considering each module separately) overlooks the big-picture interactions between the different work processes. Within our project, four principal subprocesses existed: the clinical research, the IRB, the app development, and informatics. The tight coupling of these four processes created conditions where an inability to complete one step became a rate-limiting factor for progress in other parts of the project, even in workflows that seemed unrelated (Figure 1). For example, changes in the data storage structure in the informatics workflow necessitated additional consultations with the IRB. The IRB process had to be completed before the clinical workflow could proceed. Likewise, difficulties with getting the research software installed on patient phones, due to a Wi-Fi configuration issue, required changes to the underlying software and a subsequent change to the informatics workflow.

Figure 1. Project workflows. IRB: Institutional Review Board; SFTP: secure file transfer protocol; WORDS: Women and Oncologists Reaching Decisions about Surgery.



Conclusions

The use of mHealth tools for scaling up health care research brings together new stakeholder groups, including IT teams and patients, which creates new opportunities. Like any novel research method, the use of mHealth for data collection carries a unique set of challenges for PIs who are used to traditional clinical research. The differences in pacing and work tempos between clinical research and software-driven, patient-participatory research are initially unexpected and have to be adjusted for. When combined with the variable pacing of software development, researchers should be prepared to devote additional time and energy to coordinating both the internal and external members of the research team. This time and energy will likely prove to be good investments, as mHealth techniques will likely gain prominence in the research community since they allow the clinical research community to collect valuable data in new ways. Unfortunately, the same unique qualities that allow mHealth technology to collect new types of data require adjustment in supporting infrastructures as well. While some institutions have pioneered new methods of supporting this work, many seem either unaware of new needs or unable to provide mHealth research support. Ultimately, mHealth will benefit clinical research by providing us the opportunities to gain insights into how patients actually live their lives, allowing us to create treatment strategies that work with, not against, patient lifestyles.

Conflicts of Interest

None declared.

References

1. Langford A, Orellana K, Kalinowski J, Aird C, Buderer N. Use of tablets and smartphones to support medical decision making in US adults: Cross-sectional study. JMIR Mhealth Uhealth 2020 Aug 12;8(8):e19531 [FREE Full text] [doi: 10.2196/19531] [Medline: 32784181]
2. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: A systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
3. Haghi M, Thurow K, Stoll R. Wearable devices in medical Internet of Things: Scientific research and commercially available devices. Healthc Inform Res 2017 Jan;23(1):4-15 [FREE Full text] [doi: 10.4258/hir.2017.23.1.4] [Medline: 28261526]
4. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med 2015 Apr 15;7(283):283rv3 [FREE Full text] [doi: 10.1126/scitranslmed.aaa3487] [Medline: 25877894]
5. McKeon A, McCue M, Skidmore E, Schein M, Kulzer J. Ecological momentary assessment for rehabilitation of chronic illness and disability. Disabil Rehabil 2018 Apr;40(8):974-987. [doi: 10.1080/09638288.2017.1280545] [Medline: 28637126]
6. Poelman MP, van Lenthe FJ, Scheider S, Kamphuis CB. A smartphone app combining global positioning system data and ecological momentary assessment to track individual food environment exposure, food purchases, and food consumption: Protocol for the observational FoodTrack study. JMIR Res Protoc 2020 Jan 28;9(1):e15283 [FREE Full text] [doi: 10.2196/15283] [Medline: 32012100]

7. Fitzpatrick S, Kranzler A, Fehling K, Lindqvist J, Selby EA. Investigating the role of the intensity and duration of self-injury thoughts in self-injury with ecological momentary assessment. Psychiatry Res 2020 Feb;284:112761. [doi: 10.1016/j.psychres.2020.112761] [Medline: 31951869]
8. Pryss R, Schobel J, Hoppenstedt B, Spiliopoulou M, Langguth B, Probst T, et al. Ecological momentary assessment based differences between Android and iOS users of the TrackYourHearing mHealth crowdsensing platform. Annu Int Conf IEEE Eng Med Biol Soc 2019 Jul;2019:3951-3955. [doi: 10.1109/EMBC.2019.8857854] [Medline: 31946737]
9. Pedersen A, Hack TF. Pilots of oncology health care: A concept analysis of the patient navigator role. Oncol Nurs Forum 2010 Jan;37(1):55-60. [doi: 10.1188/10.ONF.55-60] [Medline: 20044339]
10. Stead WW, Searle JR, Fessler HE, Smith JW, Shortliffe EH. Biomedical informatics: Changing what physicians need to know and how they learn. Acad Med 2011;86(4):429-434. [doi: 10.1097/acm.0b013e3181f41e8c]
11. Mobile health (mHealth) technology. OU Health. URL: https://www.ouhealth.com/stephenson-cancer-center/cancer-research/cancer-shared-resources-services/mobile-health-mhealth-technology/ [accessed 2021-11-20]
12. Wiggins A, Wilbanks J. The rise of citizen science in health and biomedical research. Am J Bioeth 2019 Aug;19(8):3-14. [doi: 10.1080/15265161.2019.1619859] [Medline: 31339831]
13. Liu Z, Ng M, Gunasekeran DV, Li H, Ponampalam K, Ponampalam R. Mobile technology: Usage and perspective of patients and caregivers presenting to a tertiary care emergency department. World J Emerg Med 2020;11(1):5-11 [FREE Full text] [doi: 10.5847/wjem.j.1920-8642.2020.01.001] [Medline: 31892997]
14. Sampat B, Prabhakar B. Privacy risks and security threats in mHealth apps. J Int Technol Inf Manag 2017;26(4):126-153 [FREE Full text]
15. mHealth Center, Penn Medicine. URL: http://mhealth.pennmedicine.org/ [accessed 2021-11-20]
16. Sharma A, Harrington RA, McClellan MB, Turakhia MP, Eapen ZJ, Steinhubl S, et al. Using digital health technology to better generate evidence and deliver evidence-based care. J Am Coll Cardiol 2018 Jun 12;71(23):2680-2690 [FREE Full text] [doi: 10.1016/j.jacc.2018.03.523] [Medline: 29880129]
17. Prokosch HU, Ganslandt T. Perspectives for medical informatics. Reusing the electronic medical record for clinical research. Methods Inf Med 2009;48(1):38-44. [Medline: 19151882]
18. CHAI Core: Interventions. UNC Gillings School of Global Public Health. URL: https://sph.unc.edu/norc/chai/ [accessed 2020-02-01]
19. Duke Mobile App Gateway. URL: https://www.mag.mobile.duke.edu/ [accessed 2020-02-01]

Abbreviations

AWS: Amazon Web Services
DMZ: demilitarized zone
IRB: Institutional Review Board
IT: information technology
mHealth: mobile health
PHI: protected health information
PI: principal investigator
SFTP: secure file transfer protocol
WORDS: Women and Oncologists Reaching Decisions about Surgery

Edited by L Buis; submitted 25.07.21; peer-reviewed by K Chen, S Hume; comments to author 24.11.21; revised version received 18.01.22; accepted 16.02.22; published 01.04.22.

Please cite as:
Chettri S, Wang V, Balkin EA, Rayo MF, Lee CN
Development of a Mobile App for Clinical Research: Challenges and Implications for Investigators
JMIR Mhealth Uhealth 2022;10(4):e32244
URL: https://mhealth.jmir.org/2022/4/e32244
doi:10.2196/32244
PMID:35363154

Viewpoint

Wearables in Schizophrenia: Update on Current and Future Clinical Applications

Lakshan N Fonseka[1], MS; Benjamin K P Woo[1], MD

Olive View-University of California Los Angeles Medical Center, Sylmar, CA, United States

Corresponding Author:
Lakshan N Fonseka, MS
Olive View-University of California Los Angeles Medical Center
14445 Olive View Drive
Sylmar, CA, 91342
United States
Phone: 1 (747) 210 3000
Email: lfonseka@ucla.edu

Abstract

Schizophrenia affects 1% of the world population and is associated with a reduction in life expectancy of 20 years. The increasing prevalence of both consumer technology and clinical-grade wearable technology offers new metrics to guide clinical decision-making remotely and in real time. Herein, recent literature is reviewed to determine the potential utility of wearables in schizophrenia, including their utility in diagnosis, first-episode psychosis, and relapse prevention and their acceptability to patients. Several studies have further confirmed the validity of various devices in their ability to track sleep—an especially useful metric in schizophrenia, as sleep disturbances may be predictive of disease onset or the acute worsening of psychotic symptoms. Through machine learning, wearable-obtained heart rate and motor activity were used to differentiate between controls and patients with schizophrenia. Wearables can capture the autonomic dysregulation that has been detected when patients are actively experiencing paranoia, hallucinations, or delusions. Multiple platforms are currently being researched, such as Health Outcomes Through Positive Engagement and Self-Empowerment, Mobile Therapeutic Attention for Treatment-Resistant Schizophrenia, and Sleepsight, that may ultimately link patient data to clinicians. The future is bright for wearables in schizophrenia, as the recent literature exemplifies their potential to offer real-time insights to guide diagnosis and management.

(JMIR Mhealth Uhealth 2022;10(4):e35600) doi:10.2196/35600

KEYWORDS

wearables; smartwatch; schizophrenia; digital phenotype; wearable; mHealth; mobile health; review; clinical application; clinical utility; clinical use; literature search; diagnosis; prevention

Introduction

Schizophrenia affects 1% of the world population and is associated with a reduction in life expectancy of 20 years [1]. The increasing prevalence of both consumer technology and clinical-grade wearable technology offers new metrics to guide clinical decision-making remotely and in real time [2-4]. These include standard measures, such as step count and sleep duration, but can expand through smartphone integration to include social activity, ambient light and noise sensing for sleep environment analysis, and others that will be discussed in this paper. Conversely, an increased amount of metrics can also be considered a disadvantage due to the challenge of translating the massive quantity of data generated from wearable devices into clinically relevant information. The solution for this may be using machine learning algorithms that are trained to identify pattern signatures and associate them with various clinical states, such as the onset of a manic or depressive episode.

The aim of this study is to review the recent literature on wearables and their potential clinical utility in schizophrenia, including their utility in first-episode psychosis (FEP) and diagnosis, relapse prevention, and patient safety. A literature search on PubMed was conducted for "((wearable) OR (smartwatch)) AND ((psychosis) OR (schizophrenia))" from 2018 to present. This time frame was selected because it allowed this paper to bridge the gap between prior systematic reviews and the most recent literature. Of the resulting 38 articles, 16 were either out of the scope of or not relevant to this paper. Thus, 22 of the 38 articles are presented in the context of the prior research they build upon, followed by a discussion on the future directions of this promising field.

Ethical Considerations

In consideration of the vulnerable individuals with schizophrenia and the profoundly personal data being obtained, it is necessary to prioritize ethical principles in all related research [5,6]. Chivilgina et al [1] highlight the most urgent concerns, such as data confidentiality, the lack of clear safety standards, and insufficient evidence on the impact of wearable technology on self-perception. Establishing ethical guidelines for digital technology use among psychiatric patients, particularly for the continuous, unobtrusive, passive data collection performed by wearables, is a basic prerequisite for patient protection.

Recent studies have evaluated the perceptions of the psychiatric patient population toward wearable devices. Dewa et al [7] found that young people with a psychiatric history had a positive perception of wearables and the expectation that continuous detection should result in immediate responses. Specifically in schizophrenia, wearables were found to be acceptable by patients and did not induce any significant paranoia [8].

Validity

Before wearables can have a significant clinical role, their measurements must first be validated. One obstacle is the wide variation in the devices used by researchers, as some studies use consumer products, such as the Fitbit (Fitbit LLC), while others opt for more standardized actigraphy measures that can offer increased accuracy at the cost of user convenience. Validating these methods specifically for sleep has been a recent focus in the literature, and the results are encouraging for most devices.

Several studies aimed to validate the sleep metrics of consumer devices, as wearables would be able to track sleep with minimal disturbance relative to polysomnography. Rookham et al [9] compared the Apple Watch (Apple Inc) to the clinically validated Philips Actiwatch Spectrum Pro (Koninklijke Philips NV) in 14 healthy adults. The study determined that the Apple Watch has an accuracy of 97% and sensitivity of 99% in identifying sleep, as well as a 79% specificity in detecting wakefulness. The watch tended to underestimate wakefulness after sleep onset by 5.74 minutes and overestimate total sleep time by 6.31 minutes. These results suggest that the Apple Watch is similar, in terms of sleep tracking capability, to the Philips actigraphy device, though future studies should include comparisons with the gold standard—polysomnography.

A 2021 article by Stucky et al [10] sought to validate the Fitbit Charge 2 (Fitbit LLC) against at-home polysomnography in 59 shift workers (police officers and paramedics). The Fitbit was found to overestimate rapid eye movement sleep latency by 29.4 minutes and wakefulness after sleep onset by 37.1 minutes. In addition, the Fitbit heart rate monitor showed limitations in detecting sudden heart rate changes due to decreased time resolution (ie, a decreased rate of gathering sleep data) when compared to polysomnography. The distribution of sleep episode durations was also different from the polysomnography results, and there were inaccuracies in sleep staging. These errors may have been due to the Fitbit proprietary algorithm, and they could be alleviated by having access to raw data that can be processed through open-source algorithms. Nonetheless, the study shows that the Fitbit can obtain reasonably accurate estimates of sleep and heart rate data.

These studies support prior literature that first demonstrated a lack of reliability with wearables, especially in the overestimation or underestimation of total sleep time and total wake time [11]. However, it was thought that these reliability issues may be partially due to the power of the studies, as they typically had less than 20 participants. Although the Stucky et al [10] study had 59 participants, sample sizes were otherwise consistently small across newer articles as well. Standardization and platforms for sharing data among researchers may have a role in solving this issue, and these will be detailed in further sections.

Wearables in Schizophrenia Care

Diagnosis and FEP

There is a bias in the literature toward feasibility studies and symptom monitoring rather than diagnosis or FEP identification [11,12]. This is likely because the process of psychiatric diagnosis typically involves an extended patient history, medication review, laboratory tests, and a period of behavioral observation [13]. However, although the use of wearables cannot replace the diagnostic process, real-time data have already proven to be promising in identifying physiological changes from baseline. Cella et al [14] studied 15 participants with FEP who wore a wrist wearable that recorded heart rate variability and electrodermal activity as proxies for distress. Participants also completed symptom self-assessments through a mobile phone app. The results showed that when distressing hallucinations and delusions were reported, electrodermal activity also significantly increased. Similarly, Schlier et al [15] found that when patients with schizophrenia were actively experiencing paranoia, alterations in the autonomic stress responses detected by wearables persisted until the paranoia subsided. No effects were found when patients reported hallucinations or intrusive thoughts. These results illustrate that the real-time detection of autonomic dysregulation may result in the earlier detection of psychiatric distress, which in turn may lead to having the clinical workup required for a diagnosis [15,16].

Additional advances are evident in a study by Reinertsen et al [17], in which changes in heart rate and locomotor activity were measured through a wearable patch in 16 patients with schizophrenia and 19 healthy controls. These patches measured signal complexity and interactions over time before the data were ultimately processed through a machine learning algorithm that allowed for perfect discrimination between controls and patients with schizophrenia.

Chen et al [18] used a novel privacy-preserving approach to using ambient noise levels as a measure of sociability. Wrist-worn audio bands recorded ambient noise over 1 week and classified signals based on the detected number of simultaneous speakers—a proxy for sociability. No speech content was analyzed to further protect privacy. Of 32 people,

there were 8 outpatients with schizophrenia or schizoaffective disorders, 11 outpatients with major depressive disorder without psychotic features, and 13 controls. The resulting social ambience measure (SAM) allowed healthy controls to be distinguished from individuals with depressive or psychotic disorders, who spent more time in isolation with corresponding lower levels of social ambience. Participants had also completed the Patient Health Questionnaire-9 and Generalized Anxiety Disorder-7 forms prior to study enrollment, and the severity of depressive and anxiety symptoms corresponded with the SAM as well. Although the sample size and duration of the study were small, the feasibility of the SAM as an objective measure of sociability was clear. Immediate identification allows for timely interventions and dynamic reassessments, and the SAM may potentially detect behavioral precursors of FEP or depressive episodes.

Wearable actigraphy data processed through a machine learning algorithm using slope entropy was shown to be capable of significantly differentiating among depression, mania, and remission in patients with bipolar disorder [19]. The concept of slope entropy is significantly outside the scope of this article, but the ability to use wearable data to distinguish these states holds promise for the identification of prodromal patterns in the future. With 1 in 5 people owning a smartwatch in the United States, the identification of significant changes from baseline, such as insomnia or autonomic dysregulation, may be performed unobtrusively [20]. Some studies suggest that patients with schizophrenia tend to report sleep abnormalities prior to disease onset, and wearable data may one day alert clinicians of patients who warrant further psychiatric investigation [21].

Management

Wearable data have the potential to be useful in guiding psychiatric management, particularly in the detection of relapse, which 80% of patients with schizophrenia are likely to encounter at least once within 5 years of FEP [22]. Lahti et al [23] provided 40 people with schizophrenia a wearable device to track activity levels and sleep and correlated these data with the severity of clinical symptoms, which was measured by scoring systems, such as the Positive and Negative Syndrome Scale (PANSS). The study was limited in terms of its results on relapse predictive value due to only 1 patient experiencing relapse. The authors stated that the relapse rate in the study was lower than a clinic's relapse rate and was maybe affected by a selection bias, as less stable patients may have refused to participate in the study. Nevertheless, the study demonstrates the feasibility and acceptability of using wearables to monitor patients with schizophrenia.

Currently underway is a similar study of 100 individuals with schizophrenia spectrum disorders who were each given a Fitbit Charge 3 (Fitbit LLC) and smartphone for free to track multiple metrics over 1 year. These metrics include sleep, physical activity, ambient light, and finger taps [24]. This appears to be the first use of a finger tap metric in schizophrenia, though typing on a keyboard has been studied previously as a proxy for cognition in patients with bipolar disorder. Decreased accuracy in finger taps may represent impaired concentration during a depressed state [25]. The finger tap metric pairs well with the small screens of wearables and is an interesting measure for assessing cognitive deficits and depressed states in patients with schizophrenia.

The study uses a digital phenotype platform called *Health Outcomes Through Positive Engagement and Self-Empowerment* (HOPES) that is based on the open-source Beiwe platform but includes smartphone data in addition to wearable devices. Previous research on the Beiwe platform showed that its detection rate for behavioral anomalies in the 2 weeks preceding a relapse was 71% higher than that in other time periods [26,27]. The first phase of HOPES entails observing participant behavior with the primary goal of developing machine learning algorithms that can predict relapse or readmission within 6 months. The second phase will include sending timely interventions in response to recognized pattern signatures, such as early warnings of relapse that may give participants the time to take steps toward preventing relapse [24,28]. Preliminary data from the wrist wearables of 21 patients over 1 week have been published and align with past correlations with PANSS scores, though these data are limited by the short time frame [29]. Once completed, the study will likely build upon its predecessor's ability to detect relapse while testing the efficacy of wearable-based interventions based on real-time data.

Other platforms are being studied as well, such as Mobile Therapeutic Attention for Treatment-Resistant Schizophrenia (m-RESIST)—a project that is currently testing the feasibility of its platform, wearable, and smartphone app combination [12,30]. Likewise, the Sleepsight platform was used with wearable sleep and activity data from 15 people with schizophrenia living in their homes. The protocol was demonstrated to be a feasible operation that was acceptable by patients [8].

Wearables offer the same mainstream health and fitness benefits to patients with schizophrenia. Bueno-Antequera et al [31] gave 82 outpatients with schizophrenia an armband sensor for 1 week to measure sedentary behavior and estimate cardiorespiratory fitness based on a 6-minute walking test. The results showed consistent relationships among sedentary behavior, increased BMI, and reduced cardiorespiratory fitness. These relationships remained significant even when controlling for age, symptom severity, and antipsychotic medication. Further, a clinical trial is currently testing wearables as part of 2 lifestyle interventions for young adults who are considered overweight or obese based on their BMIs and have a diagnosis of schizophrenia, schizoaffective disorder, bipolar disorder, or major depressive disorder [32]. The trial is expected to be completed by May 2022 and is likely to shed more light on the potential of wearables to address the lifestyle-related health issues of patients with psychiatric illness [33].

Future Directions

Wearables and machine learning have already made extensive progress, as the previously mentioned studies indicate, but several changes have been proposed to streamline the path for future research. For instance, the World Health Organization has published the Mobile Health Evaluation, Reporting and Assessment (mERA) checklist, which highlights 16 criteria to

guide experimental design and study interpretation, including user feedback, accessibility, replicability (requiring open-source code), scalability, cost assessment, and data security [34]. A surprising number of studies do not meet these criteria, and only 4 of 11 studies on mobile health and smartphone apps for schizophrenia reported on data security in a 2017 review [35]. The universal adoption of these guidelines may also make it possible for a database to be generated that would allow researchers to input and sort all contributed data by device type or patient population. By pooling data, the rate of reliability issues can be reduced and statistical power can be increased to overcome the common issue of having a small sample size.

Numerous studies have evaluated the validity and feasibility of wearables in psychiatric populations, but 2 adaptations may be beneficial for future research. First, there is variation among studies in what technology an investigated wearable is validated against, whether it be an actigraphy watch or the gold standard—polysomnography [9,10]. Although initial validations against other devices are more accessible for researchers and convenient for patients, especially in early research, studies have overall confirmed the validity of such wearables in these contexts. Thus, future validity studies should aim for comparisons between wearable technology and gold standards when possible.

As mentioned before, there is an abundance of validity, feasibility, and acceptability studies in the literature, as is characteristic of early research. Multiple consumer and clinical-grade devices have been successfully validated against each other as well as the gold standard, and studies can now shift toward an increasingly clinical emphasis. A 2019 review of smartphone sleep tracking in psychiatric populations showed that very few studies reported on improvements in sleep or mental health outcomes. Aledavood et al [36] recommended a patient-centered approach that prioritizes metrics, such as improving adherence, predicting risk and relapse, and determining whether these advances result in meaningful benefits for patients [36]. This approach shares similarities with the mERA checklist.

Lastly, there are tablet-based tests that are already in clinical use, such as the Cambridge Neuropsychological Test for identifying cognitive deficits and the Brief Assessment of Cognition in Schizophrenia, which can differentiate patients with severe mental illness from controls. The CogniSense app (Quest Diagnostics Incorporates) is likewise designed for tablets and has shown diagnostic rates that are equal to or higher than those of the Mini-Mental State Exam and the Mini-Cog Exam for cognitive impairment. Researchers, clinicians, and patients would all benefit if these exams were tailored to the small screen of the typical wearable [13,37]. Even if it were necessary to use condensed versions of these assessments, a wearable may still provide insights and improve patient participation through accessible reminders and the convenient completion of assessments from one's wrist. The results could then be sent to clinicians or researchers as a readout of all metrics at the time of the assessment.

Conclusions

Wearables in schizophrenia have made incredible progress in just the past 3 years. The foundation for wearables was further established by studies supporting the validity of various devices in their ability to track sleep, which is especially useful in schizophrenia, as sleep disturbances may be predictive of disease onset or the acute worsening of psychotic symptoms. Through machine learning, an analysis of heart rate and motor activity can be conducted to differentiate between controls and patients with schizophrenia. These and other metrics capture the autonomic dysregulation detected when patients are actively experiencing paranoia, hallucinations, or delusions. Several platforms that are currently being tested, such as HOPES, m-RESIST, and Sleepsight, may ultimately link wearable-derived patient data to clinicians. The future is bright for wearables in schizophrenia, as the recent literature exemplifies their potential to offer real-time insights to guide diagnosis and management.

Conflicts of Interest

None declared.

References

1. Chivilgina O, Elger BS, Jotterand F. Digital technologies for schizophrenia management: A descriptive review. Sci Eng Ethics 2021 Apr 09;27(2):25 [FREE Full text] [doi: 10.1007/s11948-021-00302-z] [Medline: 33835287]
2. Chong KPL, Guo JZ, Deng X, Woo BKP. Consumer perceptions of wearable technology devices: Retrospective review and analysis. JMIR Mhealth Uhealth 2020 Apr 20;8(4):e17544 [FREE Full text] [doi: 10.2196/17544] [Medline: 32310148]
3. Chong KP, Woo BK. Emerging wearable technology applications in gastroenterology: A review of the literature. World J Gastroenterol 2021 Mar 28;27(12):1149-1160 [FREE Full text] [doi: 10.3748/wjg.v27.i12.1149] [Medline: 33828391]
4. Fonseka LN, Woo BK. Consumer wearables and the integration of new objective measures in oncology: Patient and provider perspectives. JMIR Mhealth Uhealth 2021 Jul 15;9(7):e28664 [FREE Full text] [doi: 10.2196/28664] [Medline: 34264191]
5. Torous J, Keshavan M. A new window into psychosis: The rise digital phenotyping, smartphone assessment, and mobile monitoring. Schizophr Res 2018 Jul;197:67-68. [doi: 10.1016/j.schres.2018.01.005] [Medline: 29338959]
6. Nebeker C, Lagare T, Takemoto M, Lewars B, Crist K, Bloss CS, et al. Engaging research participants to inform the ethical conduct of mobile imaging, pervasive sensing, and location tracking research. Transl Behav Med 2016 Dec;6(4):577-586 [FREE Full text] [doi: 10.1007/s13142-016-0426-4] [Medline: 27688250]

7. Dewa LH, Lavelle M, Pickles K, Kalorkoti C, Jaques J, Pappa S, et al. Young adults' perceptions of using wearables, social media and other technologies to detect worsening mental health: A qualitative study. PLoS One 2019 Sep 18;14(9):e0222655. [doi: 10.1371/journal.pone.0222655] [Medline: 31532786]
8. Meyer N, Kerz M, Folarin A, Joyce DW, Jackson R, Karr C, et al. Capturing rest-activity profiles in schizophrenia using wearable and mobile technologies: Development, implementation, feasibility, and acceptability of a remote monitoring platform. JMIR Mhealth Uhealth 2018 Oct 30;6(10):e188 [FREE Full text] [doi: 10.2196/mhealth.8292] [Medline: 30377146]
9. Roomkham S, Hittle M, Cheung J, Lovell D, Mignot E, Perrin D. Sleep monitoring with the Apple Watch: comparison to a clinically validated actigraph [version 1; peer review: 2 approved with reservations, 1 not approved]. F1000Research. 2019 May 29. URL: https://f1000research.com/articles/8-754/v1 [accessed 2022-03-31]
10. Stucky B, Clark I, Azza Y, Karlen W, Achermann P, Kleim B, et al. Validation of Fitbit Charge 2 sleep and heart rate estimates against polysomnographic measures in shift workers: Naturalistic study. J Med Internet Res 2021 Oct 05;23(10):e26476 [FREE Full text] [doi: 10.2196/26476] [Medline: 34609317]
11. Guillodo E, Lemey C, Simonnet M, Walter M, Baca-García E, Masetti V, HUGOPSY Network, et al. Clinical applications of mobile health wearable-based sleep monitoring: Systematic review. JMIR Mhealth Uhealth 2020 Apr 01;8(4):e10733 [FREE Full text] [doi: 10.2196/10733] [Medline: 32234707]
12. Seppälä J, De Vita I, Jämsä T, Miettunen J, Isohanni M, Rubinstein K, M-RESIST Group, et al. Mobile phone and wearable sensor-based mHealth approaches for psychiatric disorders and symptoms: Systematic review. JMIR Ment Health 2019 Feb 20;6(2):e9819 [FREE Full text] [doi: 10.2196/mental.9819] [Medline: 30785404]
13. Hays R, Henson P, Wisniewski H, Hendel V, Vaidyam A, Torous J. Assessing cognition outside of the clinic: Smartphones and sensors for cognitive assessment across diverse psychiatric disorders. Psychiatr Clin North Am 2019 Dec;42(4):611-625 [FREE Full text] [doi: 10.1016/j.psc.2019.08.003] [Medline: 31672211]
14. Cella M, He Z, Killikelly C, Okruszek Ł, Lewis S, Wykes T. Blending active and passive digital technology methods to improve symptom monitoring in early psychosis. Early Interv Psychiatry 2019 Oct;13(5):1271-1275. [doi: 10.1111/eip.12796] [Medline: 30821079]
15. Schlier B, Krkovic K, Clamor A, Lincoln TM. Autonomic arousal during psychosis spectrum experiences: Results from a high resolution ambulatory assessment study over the course of symptom on- and offset. Schizophr Res 2019 Oct;212:163-170. [doi: 10.1016/j.schres.2019.07.046] [Medline: 31422861]
16. Cella M, Okruszek Ł, Lawrence M, Zarlenga V, He Z, Wykes T. Using wearable technology to detect the autonomic signature of illness severity in schizophrenia. Schizophr Res 2018 May;195:537-542 [FREE Full text] [doi: 10.1016/j.schres.2017.09.028] [Medline: 28986005]
17. Reinertsen E, Shashikumar SP, Shah AJ, Nemati S, Clifford GD. Multiscale network dynamics between heart rate and locomotor activity are altered in schizophrenia. Physiol Meas 2018 Oct 30;39(11):115001 [FREE Full text] [doi: 10.1088/1361-6579/aae1ed] [Medline: 30222594]
18. Chen W, Sabharwal A, Taylor E, Patel AB, Moukaddam N. Privacy-preserving social ambiance measure from free-living speech associates with chronic depressive and psychotic disorders. Front Psychiatry 2021 Aug 11;12:670020 [FREE Full text] [doi: 10.3389/fpsyt.2021.670020] [Medline: 34456760]
19. Cuesta-Frau D, Schneider J, Bakštein E, Vostatek P, Spaniel F, Novák D. Classification of actigraphy records from bipolar disorder patients using slope entropy: A feasibility study. Entropy (Basel) 2020 Dec 01;22(11):1243 [FREE Full text] [doi: 10.3390/e22111243] [Medline: 33287011]
20. Vogels EA. About one-in-five Americans use a smart watch or fitness tracker. Pew Research Center. 2020 Jan 09. URL: https://www.pewresearch.org/fact-tank/2020/01/09/about-one-in-five-americans-use-a-smart-watch-or-fitness-tracker/ [accessed 2022-03-31]
21. Kaskie RE, Graziano B, Ferrarelli F. Schizophrenia and sleep disorders: links, risks, and management challenges. Nat Sci Sleep 2017 Sep 21;9:227-239 [FREE Full text] [doi: 10.2147/NSS.S121076] [Medline: 29033618]
22. Robinson D, Woerner MG, Alvir JMJ, Bilder R, Goldman R, Geisler S, et al. Predictors of relapse following response from a first episode of schizophrenia or schizoaffective disorder. Arch Gen Psychiatry 1999 Mar;56(3):241-247. [doi: 10.1001/archpsyc.56.3.241] [Medline: 10078501]
23. Lahti AC, Wang D, Pei H, Baker S, Narayan VA. Clinical utility of wearable sensors and patient-reported surveys in patients with schizophrenia: Noninterventional, observational study. JMIR Ment Health 2021 Aug 09;8(8):e26234 [FREE Full text] [doi: 10.2196/26234] [Medline: 34383682]
24. Wang X, Vouk N, Heaukulani C, Buddhika T, Martanto W, Lee J, et al. HOPES: An integrative digital phenotyping platform for data collection, monitoring, and machine learning. J Med Internet Res 2021 Mar 15;23(3):e23984 [FREE Full text] [doi: 10.2196/23984] [Medline: 33720028]
25. Zulueta J, Piscitello A, Rasic M, Easter R, Babu P, Langenecker SA, et al. Predicting mood disturbance severity with mobile phone keystroke metadata: A BiAffect digital phenotyping study. J Med Internet Res 2018 Jul 20;20(7):e241 [FREE Full text] [doi: 10.2196/jmir.9775] [Medline: 30030209]
26. Barnett I, Torous J, Staples P, Sandoval L, Keshavan M, Onnela JP. Relapse prediction in schizophrenia through digital phenotyping: a pilot study. Neuropsychopharmacology 2018 Jul;43(8):1660-1666 [FREE Full text] [doi: 10.1038/s41386-018-0030-z] [Medline: 29511333]

27. Reinertsen E, Clifford GD. A review of physiological and behavioral monitoring with digital sensors for neuropsychiatric illnesses. Physiol Meas 2018 May 15;39(5):05TR01 [FREE Full text] [doi: 10.1088/1361-6579/aabf64] [Medline: 29671754]
28. Rashid NAA, Martanto W, Yang Z, Wang X, Heaukulani C, Vouk N, et al. Evaluating the utility of digital phenotyping to predict health outcomes in schizophrenia: protocol for the HOPE-S observational study. BMJ Open 2021 Oct 20;11(10):e046552 [FREE Full text] [doi: 10.1136/bmjopen-2020-046552] [Medline: 34670760]
29. Martanto W, Koh YY, Yang Z, Heaukulani C, Wang X, Rashid NAA, et al. Association between wrist wearable digital markers and clinical status in schizophrenia. Gen Hosp Psychiatry 2021;70:134-136. [doi: 10.1016/j.genhosppsych.2021.01.003] [Medline: 33541723]
30. Alonso-Solís A, Rubinstein K, Corripio I, Jaaskelainen E, Seppälä A, Vella VA, m-Resist group, et al. Mobile therapeutic attention for treatment-resistant schizophrenia (m-RESIST): a prospective multicentre feasibility study protocol in patients and their caregivers. BMJ Open 2018 Jul 16;8(7):e021346 [FREE Full text] [doi: 10.1136/bmjopen-2017-021346] [Medline: 30012788]
31. Bueno-Antequera J, Oviedo-Caro M, Munguía-Izquierdo D. Relationship between objectively measured sedentary behavior and health outcomes in schizophrenia patients: The PsychiActive project. Schizophr Res 2018 Jul;197:87-92. [doi: 10.1016/j.schres.2017.11.022] [Medline: 29174336]
32. Aschbrenner KA, Naslund JA, Gorin AA, Mueser KT, Scherer EA, Viron M, et al. Peer support and mobile health technology targeting obesity-related cardiovascular risk in young adults with serious mental illness: Protocol for a randomized controlled trial. Contemp Clin Trials 2018 Nov;74:97-106 [FREE Full text] [doi: 10.1016/j.cct.2018.10.005] [Medline: 30316998]
33. Aschbrenner KA. Lifestyle intervention for young adults with serious mental illness. ClinicalTrials.gov. URL: https://clinicaltrials.gov/ct2/show/NCT02815813 [accessed 2022-03-31]
34. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016 Mar 17;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]
35. Torous J, Firth J, Mueller N, Onnela JP, Baker JT. Methodology and reporting of mobile heath and smartphone application studies for schizophrenia. Harv Rev Psychiatry 2017;25(3):146-154 [FREE Full text] [doi: 10.1097/HRP.0000000000000133] [Medline: 28234658]
36. Aledavood T, Torous J, Hoyos AMT, Naslund JA, Onnela JP, Keshavan M. Smartphone-based tracking of sleep in depression, anxiety, and psychotic disorders. Curr Psychiatry Rep 2019 Jun 04;21(7):49 [FREE Full text] [doi: 10.1007/s11920-019-1043-y] [Medline: 31161412]
37. CogniSense on the App Store. App Store. URL: https://apps.apple.com/us/app/cognisense/id971776474 [accessed 2021-11-16]

Abbreviations

FEP: first-episode psychosis
HOPES: Health Outcomes Through Positive Engagement and Self-Empowerment
m-RESIST: Mobile Therapeutic Attention for Treatment-Resistant Schizophrenia
mERA: Mobile Health Evaluation, Reporting and Assessment
PANSS: Positive and Negative Syndrome Scale
SAM: social ambience measure

Edited by L Buis; submitted 10.12.21; peer-reviewed by B Johnson, T Dunnsiri, S Chung, S Shu; comments to author 27.01.22; revised version received 07.02.22; accepted 22.03.22; published 07.04.22.

Please cite as:
Fonseka LN, Woo BKP
Wearables in Schizophrenia: Update on Current and Future Clinical Applications
JMIR Mhealth Uhealth 2022;10(4):e35600
URL: https://mhealth.jmir.org/2022/4/e35600
doi:10.2196/35600
PMID:35389361

Viewpoint

Demographic Imbalances Resulting From the Bring-Your-Own-Device Study Design

Peter Jaeho Cho[1], BA; Jaehan Yi[1], BSc; Ethan Ho[1]; Md Mobashir Hasan Shandhi[1], BSc, MSc, PhD; Yen Dinh[1], BSc; Aneesh Patil[1]; Leatrice Martin[2], BA, MPA, MBA; Geetika Singh[1], BEng, MSc; Brinnae Bent[1], BSc, MSc, PhD; Geoffrey Ginsburg[3], BSc, MD, PhD; Matthew Smuck[4], BSc, MD; Christopher Woods[5], BA, MD, MPA; Ryan Shaw[6], BSc, MSc, PhD; Jessilyn Dunn[1,7], BSc, MSc, PhD

[1]Department of Biomedical Engineering, Duke University, Durham, NC, United States
[2]Clinical and Translational Science Institute, Duke University, Durham, NC, United States
[3]All of Us Research Program, National Institutes of Health, Bethesda, MD, United States
[4]Department of Physical Medicine and Rehabilitation, The Spine Program, University of Michigan, Ann Arbor, MI, United States
[5]Division of Infectious Diseases, Duke University Medical Center, Durham, NC, United States
[6]School of Nursing, Duke University, Durham, NC, United States
[7]Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, United States

Corresponding Author:
Jessilyn Dunn, BSc, MSc, PhD
Department of Biomedical Engineering
Duke University
Room 1427, Fitzpatrick Center (FCIEMAS)
101 Science Drive
Durham, NC, 27708-0281
United States
Phone: 1 919 660 5131
Email: jessilyn.dunn@duke.edu

Abstract

Digital health technologies, such as smartphones and wearable devices, promise to revolutionize disease prevention, detection, and treatment. Recently, there has been a surge of digital health studies where data are collected through a bring-your-own-device (BYOD) approach, in which participants who already own a specific technology may voluntarily sign up for the study and provide their digital health data. BYOD study design accelerates the collection of data from a larger number of participants than cohort design; this is possible because researchers are not limited in the study population size based on the number of devices afforded by their budget or the number of people familiar with the technology. However, the BYOD study design may not support the collection of data from a representative random sample of the target population where digital health technologies are intended to be deployed. This may result in biased study results and biased downstream technology development, as has occurred in other fields. In this viewpoint paper, we describe demographic imbalances discovered in existing BYOD studies, including our own, and we propose the Demographic Improvement Guideline to address these imbalances.

(JMIR Mhealth Uhealth 2022;10(4):e29510) doi:10.2196/29510

KEYWORDS

bring your own device; wearable device; mHealth

Introduction

Digital health tools, including mobile health (mHealth) and wearable devices, can provide researchers with high-frequency data that are more representative of a person's health state during their daily life than what can be collected in a clinical setting [1,2]. The enormous benefits of acquiring data outside of the clinic have led researchers to adopt new study designs to incorporate digital health data collection tools. Digital biomarkers constitute a type of biomarker that is developed from digitally collected data, such as wearable devices and smartphones, to evaluate functions and processes in the body and that can typically be recorded outside of a lab setting to provide continuous and more holistic information [3]. In

particular, the bring-your-own-device (BYOD) study design has become increasingly popular because it gives researchers the ability to collect large-scale data at a low cost from participants who already own personal electronic devices, such as smartphones and wearable devices. During the COVID-19 pandemic, there has been growing interest in using digital health data to track illness, either for COVID-19 detection or to support telemedicine [4-6]. To process all this data, artificial intelligence algorithms, specifically machine learning algorithms, are being developed to detect health conditions by learning from previously collected data.

Machine learning algorithms rely on data to train models; they are susceptible to biases that result in poor predictions for segments of the population if the training data are not representative of the population where the model is intended to be used [7]. Therefore, one key aspect of machine learning is the data collection process, whereby researchers identify the target population and select a representative, random sample of the population from which to collect unbiased data [8]. BYOD studies are particularly susceptible to bias in the data collection process because the recruitment pool is limited to people who already own a device, and this population is generally not the only one where the tools are ultimately intended to be used in practice. In BYOD studies, a nonrepresentative study population excludes key socioeconomic and physiologic circumstances that can covary with race, ethnicity, or both. For example, disease prevalence and pathophysiology often vary by race and ethnicity (eg, COVID-19 infection and mortality rates [9-11], manifestation of metabolic disease [12-14], cardiovascular disease [15,16], and sleep irregularities [17,18]), which can result in differences in how the newly developed technologies will perform. As a specific example, optical sensors to measure blood oxygen saturation may fail in people with more melanin and for people with the sickle cell trait [19-21], both characteristics common in Black populations, for two completely separate reasons. To address such problems, representative digital health data are needed for building generalizable machine learning models that are as accurate under deployment as they are in the initial testing phase. Like other fields that have discovered that bias in data used to train models has led to biased models, we fear that digital health will face similar challenges if the bias inherent in BYOD studies is left unaddressed [22,23]. In this viewpoint paper, we seek to raise awareness of demographic imbalances in several BYOD studies and propose a guideline based on our own BYOD case study to directly improve the demographic balance of BYOD digital health studies.

Demographic Imbalances in Existing BYOD Studies

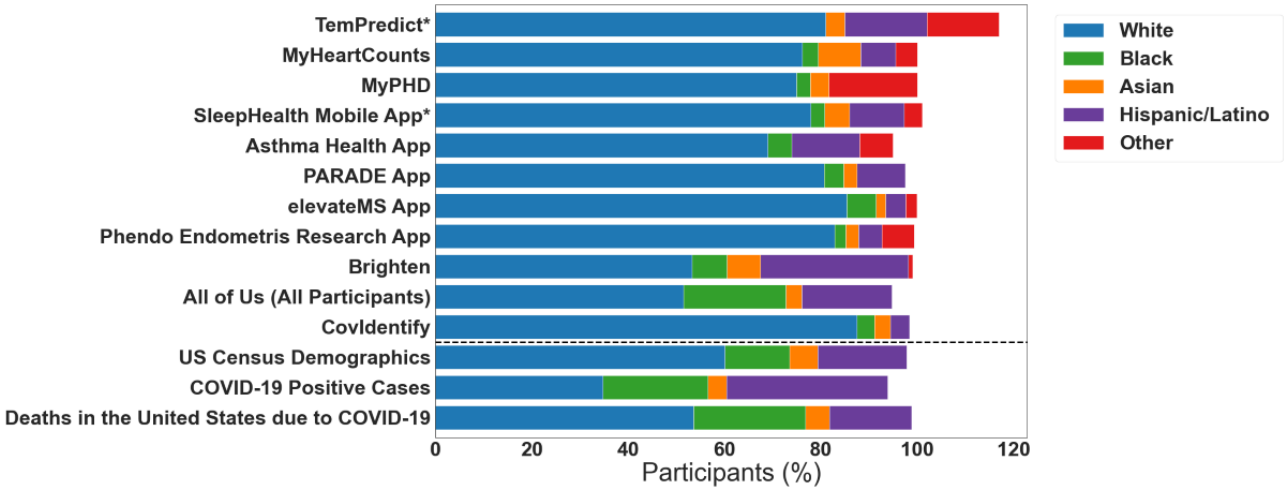
BYOD is a term used to describe studies in which participants contribute data from self-owned personal electronic devices. We collected a sample of BYOD studies using multiple search terms on PubMed and Web of Science, including “bring your own device,” “(consumer) wearable device study,” “remote mobile study,” and “mHealth,” and performed manual review and filtering, which generated 15 relevant studies. Although we note that this is not a systematic approach to compiling all published BYOD studies, and we acknowledge the potential to overlook studies that have successfully recruited a representative study population, here we demonstrate that many existing BYOD studies have gender and race imbalances when compared to the broader US demographic profile. Of the 15 studies identified, 4 (27%) did not report any demographics on ethnicity or race, and none of the remaining 11 (73%) studies achieved participant demographic proportions representative of the overall US demographic profile (Multimedia Appendix 1).

One of the most pre-eminent BYOD studies was a substudy of the All of Us research initiative [24] by the US National Institutes of Health (NIH). In that substudy, wearable device data were collected from participants who owned Fitbit devices between 2008 and 2019 and who consented to share their data. More than 80% of participants in the overarching All of Us study were from historically underrepresented groups in biomedical research. As a strong juxtaposition, 70% of the participants in the All of Us BYOD substudy identified as White non-Hispanic, while only 4% and 3% identified as Black and Asian, respectively (Figure 1 and Multimedia Appendix 2). The ethnicity distribution in the All of Us study tells a similar story, with over 90% of the participants identifying as non-Hispanic or non-Latino, and only 6% identifying as Hispanic or Latino. Even in this diverse, large-scale study that had specifically targeted recruitment toward underrepresented groups, equitable demographic representation was limited by the BYOD study design.

Similar to other BYOD studies, we discovered demographic imbalances in our own CovIdentify study (Multimedia Appendices 3 and 4). The unique circumstances of the evolving COVID-19 pandemic led to a rapid launch of our study, where we aimed to develop an intelligent testing strategy using digital biomarkers extracted from personally owned commercial wearable devices under resource-constrained conditions (limited testing, rural areas, etc). However, after our rapid study launch, we found that the communities most vulnerable to COVID-19 [9,25] were not well-represented in our study.

To mitigate this demographic imbalance and to ensure the inclusion of participants from underserved communities, we developed the Demographic Improvement Guideline and correspondingly altered our recruitment process. Although we developed this guideline retrospectively after the discovery of demographic imbalances in our study, we are calling for future research to take proactive measures during the BYOD study design as well as responsive measures during participant recruitment and retention efforts. Many BYOD studies acknowledge demographic imbalance as a limitation, and we believe a concerted effort is needed to enact change to reduce bias in digital health data used in research.

Figure 1. Comparing demographic distributions from various bring-your-own-device (BYOD) studies (listed on the y-axis, above the dotted line), the national census, and the National Vital Statistics System. Studies with an asterisk did not separate ethnicity and race and, therefore, have percentages that sum up to be greater than 100. Other studies did not report a breakdown for all the ethnicity and race groups and, therefore, resulted in an aggregated percentage less than 100.



Demographic Improvement Guideline

Overview

The goal of BYOD studies is to develop new device-based technologies and interventions to improve the health or well-being of populations. In order for these interventions to be fit for purpose, the research and development should include the populations where the technologies will ultimately be used, or their exclusion should be well-justified and should not introduce bias or harm. There may be cases where demographic imbalances (ie, sampling bias) would not be problematic in a BYOD study, for example, if both of the following conditions apply: (1) the manifestation of the disease as measured by the wearable device does not differ across race, ethnicity, or age and (2) the technology works the same for all people across the entire population. In such cases, the researcher would not need to focus efforts and resources on obtaining a representative demographic distribution in the study population.

However, it is known that (1) most diseases do not manifest in the same way across different demographics and (2) smart devices do not work equally well or are not equally accessible across all demographics. As such, we cannot conclude that a biased sampling strategy is generally an appropriate choice. There will certainly be exceptions where the study population is appropriately focused on a certain group (eg, only including females in pregnancy-related studies), which can be achieved by applying inclusion and exclusion criteria for study participation. In this case, before a biased sampling strategy is chosen, the researcher is obligated to prove the null hypothesis for differences in disease manifestation and device function between the biased sample and the population where the technology is intended to be used. We purport that this step is often an even larger barrier than designing an equitably sampled study population. Next, we describe two concrete examples of times when equitable sampling was not prioritized and resulted in incorrect study conclusions.

The first condition for considering sampling bias is when domain knowledge from a field gives a priori indication that disease prevalence and pathophysiology may vary by race, ethnicity, or both, for example, COVID-19 infection and mortality rates, metabolic disease, cardiovascular disease, and sleep disturbances, among others [9-18]. Coronary artery disease is a historical example of the unintended and harmful effects that biased study populations can have on study conclusions: most early coronary artery disease studies consisted of homogenous male populations and, as a result, differences in symptoms for men and women were not discovered until follow-up studies included female populations [26,27]. Often there is insufficient a priori knowledge of differences in disease manifestation across age, gender, race, ethnicity, etc. As a result, the effect of biased sampling on study conclusions is frequently unknown. Furthermore, data quality and sampling often vary across demographics, particularly in mobile and wearable device studies, which further complicates determining the most appropriate sampling method. For example, use of wearable devices, ranging from commercial wearable devices to more sophisticated health monitoring devices, is lower among adults over 50 years of age compared to adults aged 18 to 34 years. Young, healthy, and more educated individuals are more likely to own wearable devices [28].

The second condition for considering sampling bias is when the technology used for data collection has not been validated systematically (ie, it has either not reported demographic distributions of the test population or has uncertainty in its measurements). For example, a study published in 2020 in the New England Journal of Medicine compared values of blood oxygen saturation in occult hypoxia measured via pulse oximetry with arterial oxygen saturation in arterial blood gas, the gold standard, to determine the validity of the pulse oximetry measurements [19]. The study obtained nearly 50,000 pairs of measurements from more than 8000 White patients and 1000 Black patients. The frequency of occult hypoxia detection via pulse oximetry was three times lower for Black patients than White patients. Given the prevalent use of pulse oximetry in medical decision-making, the implication of sensors reporting varying results based on an individual's skin tone is concerning. The study points to the need for manufacturers of optical heart

rate and blood oxygen saturation sensors to disclose the demographic distribution of the populations that were used to test the sensors. Since the primary sensor on most of the wrist-worn commercial wearable devices (eg, Apple Watch, Fitbit, and Garmin) relies on a similar optical sensor, ensuring that the measurements are accurate for anyone who wears the device is crucial. Because the performance of technology across demographic characteristics is not systematically evaluated and published for most commercial wearable devices, the technology and models derived from them may fail to generalize across demographic characteristics [20,29].

This guideline is relevant to studies in which sampling bias resulting in demographic imbalance could challenge the validity and generalizability of a BYOD study's conclusions. The method can be implemented iteratively in the study design and execution process, and includes the following steps:

1. Identify one or more populations at risk of being omitted from the study for whom the technology may ultimately be used and determine if BYOD study design is appropriate for the research question.
2. If the BYOD study design is insufficient for addressing issues associated with demographic imbalance, modify the study design using internal and external resources to improve dissemination of information and improve engagement with the target populations.
3. Launch the study and monitor study demographics in real time to adjust downstream efforts accordingly.

Identify Populations of Interest

To identify the populations that are at risk of being omitted, a literature review can reveal a baseline expectation of demographic distributions from prior studies using similar devices and advertisement strategies. To support this, it is necessary for publication venues and funding agencies to require detailed demographic reporting of BYOD studies. In addition to the proactive measures mentioned above, researchers should conduct early and ongoing systematic analyses of their study demographics and iteratively adapt their strategies accordingly.

Modify Study Design

Capabilities to disseminate information, provide physical resources, and improve engagement with the target population can be assessed and developed internally and externally. Internal resources may involve organizations within the research institution that have experience recruiting underrepresented populations or have ties with the underrepresented groups. Another internal approach may be choosing the devices to be included in the study, considering whether these devices have widespread use in the underrepresented groups, and augmenting and using capabilities of available devices (eg, using sensors of more widely available smartphones to capture physiology and activity characteristics instead of consumer wearables) [30]. External resources may include community groups that are representative of the target population, government and nongovernmental organizations that work with the underrepresented communities, or donors who can donate devices that can be distributed to the target population. Other external resources may include clinician referral, which can help establish trust with the target population and improve retention of study participants [31].

For national or international studies, researchers can partner with nationwide organizations, launch social media advertising campaigns, or both [32-34]. For regional studies, researchers can connect with institutional and local resources to learn about and connect with community groups. Researchers may recruit a liaison to aid in establishing partnerships with external organizations and establish a community advisory board to interact directly with community groups to contribute to the study design and advocate for participation in the study.

One existing challenge is a dearth of funding to support equitable digital health study design, including the purchase of personal electronic devices, such as wearables and smartphones. Funders should be aware of this challenge and develop funding opportunities to support equitable digital health research. Given that acquiring funding is a well-known challenge, there are methods by which researchers can alter study design to reduce costs. This can include altering the study to follow a subset of the study population based on specific cases and controls. One method is the implementation of a randomized withdrawal design where participants who are inactive at the beginning of the study are excluded so that participants who are active are presented in the actual randomized study [35]. This could limit the number of participants that researchers need to follow up on and provide a more robust data set, but this may also skew demographics. Another challenge is to ensure that the distributed devices are used as intended, given that they require a certain level of digital literacy and need to be consistently worn and charged over time. As such, participant compensation amount and timing as well as wear tracking should be included as considerations in the study design and should be supported by funding agencies [36].

Launch Study and Reassess Study Population

After study modifications, the study is ready for deployment. Depending on the funding received and study design changes, the resulting study may be a hybrid study where devices are provided for the target population. During this launch phase, researchers should regularly assess their demographic distributions to determine whether their strategy is effective. If the strategy is not effective, researchers can restrategize with their internal and external community partners and identify aspects of the study that may be deterring participant engagement or recruitment. Finally, study teams should continue to engage with the study population after the end of the study to convey findings and future opportunities for participation in research.

A Case Study on Practically Implementing the Demographic Improvement Guideline

Overview

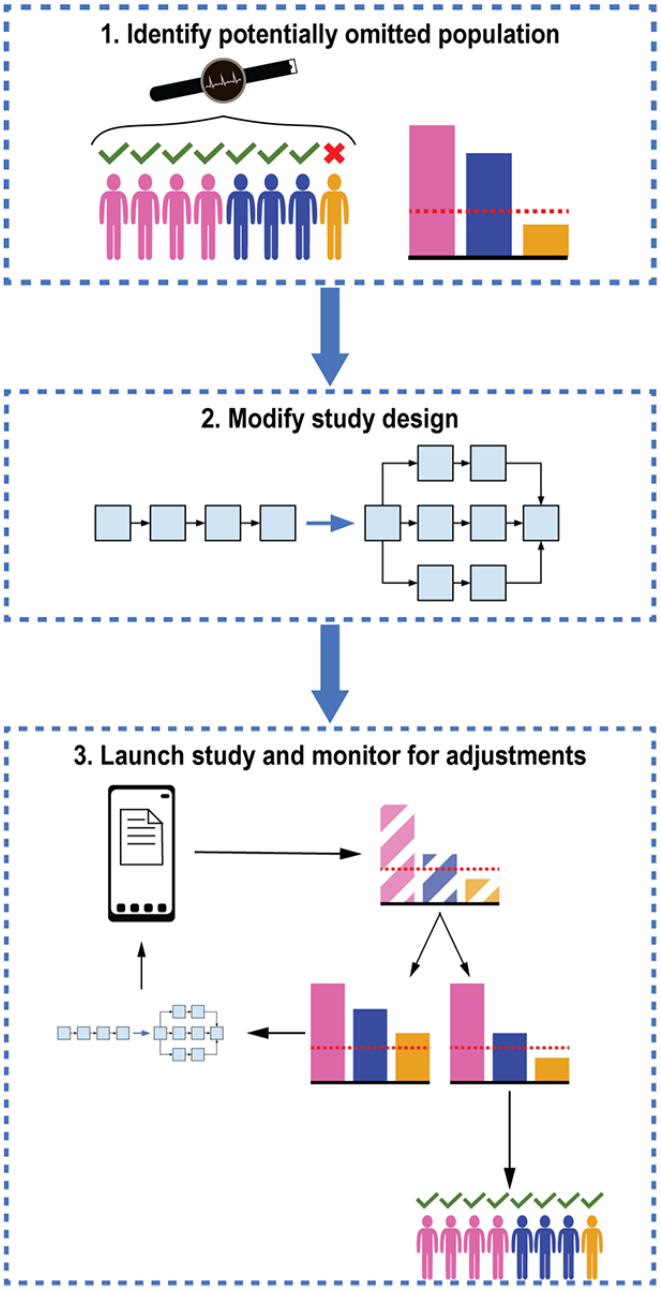
The following subsections provide details about the responsive implementation of the Demographic Improvement Guideline after the launch of the CovIdentify study (Institutional Review Board No. 2020-0412). The study aimed to develop machine learning algorithms to detect COVID-19 and influenza from

wearable device data, with a long-term vision of developing an intelligent diagnostic testing strategy using digital biomarkers extracted from personally owned commercial wearable devices under resource-constrained conditions (limited testing, rural areas, etc). In April 2020, CovIdentify began enrolling participants. Following informed consent, participants contributed their wearable device data (eg, Fitbit, Garmin, and Apple Watch) and reported daily symptoms for 12 months via a downloadable app, email, or text message.

Identify

Following the rapid launch, exploratory data analysis revealed major differences between CovIdentify demographics and the demographics of COVID-19–positive cases and deaths in the United States [30], as well as overall US demographics based on the 2020 US Census. The communities hardest hit by the COVID-19 pandemic, including Black and African American as well as Hispanic and Latinx communities, also had the lowest representation (Figure 2) [9,37,38].

Figure 2. Visualization of the Demographic Improvement Guideline. Step 1. Identify the populations at risk of being omitted from the study and for whom the technology may ultimately be used. Step 2. Modify study design based on internal and external resources to disseminate information and improve engagement with the target populations. Step 3. Launch the study, monitor study demographics in real time, and adjust downstream efforts accordingly. Researchers should reassess their study population demographics to ensure that target distributions are achieved and restrategize accordingly. The red dashed lines in the bar charts are visual representations of the acceptable threshold for each subgroup population.



Modify

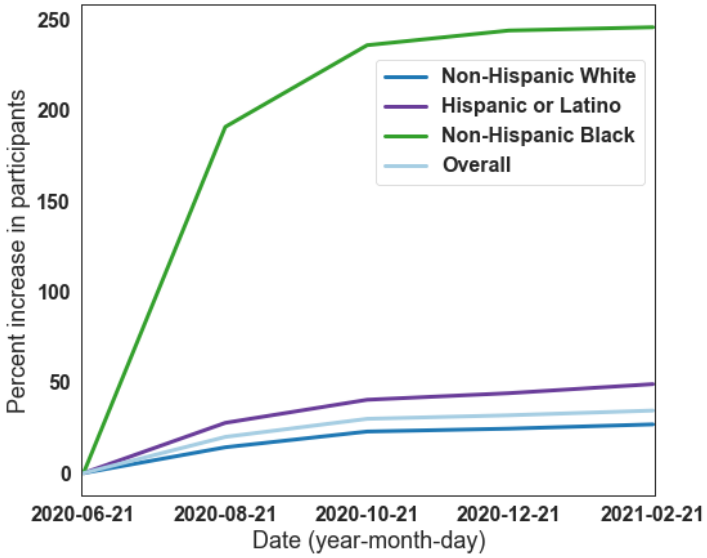
To address this imbalance, we designed the Demographic Improvement Guideline mitigation strategy in partnership with Duke University's Clinical and Translational Science Institute (CTSI), an NIH-funded center that connects researchers with local community partners and improves the reach and efficacy of research [39]. We piloted this method in the Durham, North

Carolina community from June through October of 2020. The CTSI facilitated connections between our research team and local community and faith leaders, including groups working to address COVID-19–related health disparities affecting the Latinx and African American communities. This engagement enabled our team to learn directly from community members, researchers, and health care professionals in order to improve our advertisement and recruitment strategies to (1) increase awareness about the study and (2) distribute wearable devices that were donated to our study and purchased through research funding. To expand awareness of the study, our team gave presentations to community groups regarding the advantages of continuous health monitoring, the uses of participant information in this study, and imbalances in our current study population that would make it difficult for our team to develop generalizable study findings. We also recruited a liaison to the Latinx community, translated our website to four additional languages, ran multilingual social media advertisements featuring diverse images and videos, and posted messages about the study on various social media platforms.

Launch Study and Reassess Study Population

To support the purchase of devices and social media advertising campaigns, we applied for nearly 30 funding opportunities from government, nonprofit, and industry sources. We were awarded a Duke Bass Connections Fellowship, a North Carolina Biotechnology Institute grant, and a Duke MEDx/CTSI award that enabled us to purchase 65 wearable devices for distribution. We also received a donation of 300 additional devices from a previously completed study. We attended 12 community events, including food and medication distribution events for low-income members of the Durham community, and distributed 250 free wearable devices in a socially distant manner to ensure safety during the COVID-19 pandemic. It should be noted that this was a hybrid BYOD design because participants were still required to own a smartphone to connect their wearable device to the study. We also worked with wearables companies to set up reduced device pricing with a direct link from our study's main webpage to improve accessibility. Together, these efforts resulted in a 250% increase in the representation of Black and African American participants and a 49% increase in the Latinx and Hispanic population within 4 months of the implementation of the guideline (Figure 3).

Figure 3. Percent increase in CovIdentify study population participants compared to June 21, 2020.



Discussion

In this viewpoint paper, we discussed the need for digital health studies to sample from populations representative of the target population to ensure equitable performance of predictive or machine learning models. We explored demographic imbalances in BYOD studies and proposed the Demographic Improvement Guideline to address these imbalances with an implementation example from the CovIdentify study. We believe that future efforts and funding in this space can further improve equitable digital health study design and data collection. Further, we recommend that researchers carefully consider the financial incentives and personal motivations provided to participants by the study to identify driving factors for participation and engagement.

By developing the Demographic Improvement Guideline, we aim to facilitate improvement of future BYOD study design through fostering relationships and trust with local community groups. These methods are not limited to BYOD studies only. We can translate these methods to non-BYOD studies as well. In addition, we believe that implementing existing community-based engagement methods, such as training recruiters and providing face-to-face screening, can improve both recruitment methods and adherence to studies [40-43]. We also emphasize the need for increased funding opportunities in this area to enable the development of equitable algorithms and models that are representative of all individuals.

One limitation of the Demographic Improvement Guideline was the condition under which it was developed. While the guideline is a proactive recommendation, the implementation within our case study was retrospective to our findings.

Therefore, shifts in our study design, such as donating commercial wearable devices to underrepresented groups, have resulted in a non-BYOD (ie, hybrid) study. Furthermore, due to the evolving nature of the pandemic and the resulting rapid launch of the CovIdentify study, factors mentioned in the guideline may not be applicable to all study designs. Here, we intend to present the guideline as one unique potential solution of the many possible solutions to address demographic imbalances in BYOD studies. This experience underscores the importance of addressing potential demographic inequities prior to a BYOD study launch.

Acknowledgments

This research was supported by the Bass Connections Fellowship; the Duke–Robert J. Margolis, MD, Center for Health Policy at Duke University; MEDx; NCBiotech awards; and the Duke CTSI Community Engaged Research Initiative. We would like to acknowledge Rosa Gonzalez-Guarda, Leonor Corsino, Sabrena Mervin-Blake, Nicholas Eberlein, Jamie Roberts, and Kelly Keefe. Support was also provided, in part, by Duke University's CTSI by a US NIH Clinical and Translational Science Award (grant UL1TR002553). This article was prepared while GG was employed at Duke University. The opinions expressed in this article are the authors' own and do not reflect the views of the NIH, the Department of Health and Human Services, or the United States government.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Race and ethnicity breakdown from bring-your-own-device (BYOD) studies.
[DOCX File , 33 KB - mhealth_v10i4e29510_app1.docx]

Multimedia Appendix 2
Sex, race, and ethnicity reported by various bring-your-own-device (BYOD) studies, the national census, and the National Vital Statistics System.
[DOCX File , 22 KB - mhealth_v10i4e29510_app2.docx]

Multimedia Appendix 3
Wearable device ownership by race or ethnicity for CovIdentify case study.
[DOCX File , 125 KB - mhealth_v10i4e29510_app3.docx]

Multimedia Appendix 4
Wearable device ownership of iPhone (iOS) users by race or ethnicity for CovIdentify case study.
[DOCX File , 125 KB - mhealth_v10i4e29510_app4.docx]

References

1. Nelson BW, Low CA, Jacobson N, Areán P, Torous J, Allen NB. Guidelines for wrist-worn consumer wearable assessment of heart rate in biobehavioral research. NPJ Digit Med 2020;3:90 [FREE Full text] [doi: 10.1038/s41746-020-0297-4] [Medline: 32613085]
2. Dunn J, Kidzinski L, Runge R, Witt D, Hicks JL, Schüssler-Fiorenza Rose SM, et al. Wearable sensors enable personalized predictions of clinical laboratory measurements. Nat Med 2021 Jun;27(6):1105-1112 [FREE Full text] [doi: 10.1038/s41591-021-01339-0] [Medline: 34031607]
3. Coravos A, Khozin S, Mandl KD. Developing and adopting safe and effective digital biomarkers to improve patient outcomes. NPJ Digit Med 2019;2(1):14 [FREE Full text] [doi: 10.1038/s41746-019-0090-4] [Medline: 30868107]
4. Baumgart DC. Digital advantage in the COVID-19 response: Perspective from Canada's largest integrated digitalized healthcare system. NPJ Digit Med 2020;3:114 [FREE Full text] [doi: 10.1038/s41746-020-00326-y] [Medline: 32923691]
5. Li X, Dunn J, Salins D, Zhou G, Zhou W, Schüssler-Fiorenza RSM, et al. Digital health: Tracking physiomes and activity using wearable biosensors reveals useful health-related information. PLoS Biol 2017 Jan;15(1):e2001402 [FREE Full text] [doi: 10.1371/journal.pbio.2001402] [Medline: 28081144]
6. Radin JM, Wineinger NE, Topol EJ, Steinhubl SR. Harnessing wearable device data to improve state-level real-time surveillance of influenza-like illness in the USA: A population-based study. Lancet Digit Health 2020 Feb;2(2):e85-e93 [FREE Full text] [doi: 10.1016/S2589-7500(19)30222-5] [Medline: 33334565]
7. Panch T, Mattie H, Atun R. Artificial intelligence and algorithmic bias: Implications for health systems. J Glob Health 2019 Dec;9(2):010318 [FREE Full text] [doi: 10.7189/jogh.09.020318] [Medline: 31788229]

8. Paulus JK, Kent DM. Predictably unequal: Understanding and addressing concerns that algorithmic clinical prediction may increase health disparities. NPJ Digit Med 2020;3:99 [FREE Full text] [doi: 10.1038/s41746-020-0304-9] [Medline: 32821854]
9. COVID-19. US Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/index.html [accessed 2020-06-14]
10. Gray DM, Anyane-Yeboa A, Balzora S, Issaka RB, May FP. COVID-19 and the other pandemic: Populations made vulnerable by systemic inequity. Nat Rev Gastroenterol Hepatol 2020 Sep;17(9):520-522 [FREE Full text] [doi: 10.1038/s41575-020-0330-8] [Medline: 32541960]
11. Shah M, Sachdeva M, Dodiuk-Gad RP. COVID-19 and racial disparities. J Am Acad Dermatol 2020 Jul;83(1):e35 [FREE Full text] [doi: 10.1016/j.jaad.2020.04.046] [Medline: 32305444]
12. Lacy ME, Wellenius GA, Sumner AE, Correa A, Carnethon MR, Liem RI, et al. Association of sickle cell trait with hemoglobin A1c in African Americans. JAMA 2017 Feb 07;317(5):507-515 [FREE Full text] [doi: 10.1001/jama.2016.21035] [Medline: 28170479]
13. Ziemer DC, Kolm P, Weintraub WS, Vaccarino V, Rhee MK, Twombly JG, et al. Glucose-independent, Black-White differences in hemoglobin A1c levels: A cross-sectional analysis of 2 studies. Ann Intern Med 2010 Jun 15;152(12):770-777. [doi: 10.7326/0003-4819-152-12-201006150-00004] [Medline: 20547905]
14. Wheeler E, Leong A, Liu CT, Hivert MF, Strawbridge RJ, Podmore C, EPIC-CVD Consortium, EPIC-InterAct Consortium, Lifelines Cohort Study, et al. Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. PLoS Med 2017 Oct;14(9):e1002383 [FREE Full text] [doi: 10.1371/journal.pmed.1002383] [Medline: 28898252]
15. Gu A, Yue Y, Desai RP, Argulian E. Racial and ethnic differences in antihypertensive medication use and blood pressure control among US adults with hypertension: The National Health and Nutrition Examination Survey, 2003 to 2012. Circ Cardiovasc Qual Outcomes 2017 Jan;10(1):e003166. [doi: 10.1161/CIRCOUTCOMES.116.003166] [Medline: 28096206]
16. Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2020 Update: A report from the American Heart Association. Circulation 2020 Mar 03;141(9):e139-e596 [FREE Full text] [doi: 10.1161/CIR.0000000000000757] [Medline: 31992061]
17. Chattu VK, Chattu SK, Spence DW, Manzar MD, Burman D, Pandi-Perumal SR. Do disparities in sleep duration among racial and ethnic minorities contribute to differences in disease prevalence? J Racial Ethn Health Disparities 2019 Dec;6(6):1053-1061. [doi: 10.1007/s40615-019-00607-7] [Medline: 31264064]
18. Chen X, Wang R, Zee P, Lutsey PL, Javaheri S, Alcántara C, et al. Racial/ethnic differences in sleep disturbances: The Multi-Ethnic Study of Atherosclerosis (MESA). Sleep 2015 Jun 01;38(6):877-888 [FREE Full text] [doi: 10.5665/sleep.4732] [Medline: 25409106]
19. Sjoding MW, Dickson RP, Iwashyna TJ, Gay SE, Valley TS. Racial Bias in Pulse Oximetry Measurement. N Engl J Med 2020 Dec 17;383(25):2477-2478 [FREE Full text] [doi: 10.1056/NEJMc2029240] [Medline: 33326721]
20. Bent B, Goldstein BA, Kibbe WA, Dunn JP. Investigating sources of inaccuracy in wearable optical heart rate sensors. NPJ Digit Med 2020;3:18 [FREE Full text] [doi: 10.1038/s41746-020-0226-6] [Medline: 32047863]
21. Blaisdell CJ, Goodman S, Clark K, Casella JF, Loughlin GM. Pulse oximetry is a poor predictor of hypoxemia in stable children with sickle cell disease. Arch Pediatr Adolesc Med 2000 Sep;154(9):900-903. [doi: 10.1001/archpedi.154.9.900] [Medline: 10980793]
22. Panch T, Szolovits P, Atun R. Artificial intelligence, machine learning and health systems. J Glob Health 2018 Dec;8(2):020303 [FREE Full text] [doi: 10.7189/jogh.08.020303] [Medline: 30405904]
23. Sim I. Mobile Devices and Health. N Engl J Med 2019 Sep 05;381(10):956-968. [doi: 10.1056/NEJMra1806949] [Medline: 31483966]
24. All of Us research program expands data collection efforts with Fitbit. National Institutes of Health. 2019 Jan 16. URL: https://allofus.nih.gov/news-events-and-media/announcements/all-us-research-program-expands-data-collection-efforts-fitbit [accessed 2021-03-23]
25. COVID-19 in racial and ethnic minority groups. US Centers for Disease Control and Prevention. 2020. URL: https://stacks.cdc.gov/view/cdc/89820/cdc_89820_DS1.pdf? [accessed 2022-01-27]
26. Maas A, Appelman Y. Gender differences in coronary heart disease. Neth Heart J 2010 Dec;18(12):598-602 [FREE Full text] [doi: 10.1007/s12471-010-0841-y] [Medline: 21301622]
27. Kyker KA, Limacher MC. Gender differences in the presentation and symptoms of coronary artery disease. Curr Womens Health Rep 2002 May;2(2):115-119. [Medline: 12116600]
28. Chandrasekaran R, Katthula V, Moustakas E. Patterns of use and key predictors for the use of wearable health care devices by US adults: Insights from a national survey. J Med Internet Res 2020 Oct 16;22(10):e22443 [FREE Full text] [doi: 10.2196/22443] [Medline: 33064083]
29. Goldsack JC, Coravos A, Bakker JP, Bent B, Dowling AV, Fitzer-Attas C, et al. Verification, analytical validation, and clinical validation (V3): The foundation of determining fit-for-purpose for biometric monitoring technologies (BioMeTs). NPJ Digit Med 2020;3:55 [FREE Full text] [doi: 10.1038/s41746-020-0260-4] [Medline: 32337371]

30. Althoff T, Sosič R, Hicks JL, King AC, Delp SL, Leskovec J. Large-scale physical activity data reveal worldwide activity inequality. Nature 2017 Jul 20;547(7663):336-339. [doi: 10.1038/nature23018] [Medline: 28693034]
31. Funk C, Johnson C, Hefferon M. 5 key findings about public trust in scientists in the US. Pew Research Center. 2019 Aug 05. URL: https://www.pewresearch.org/fact-tank/2019/08/05/5-key-findings-about-public-trust-in-scientists-in-the-u-s/ [accessed 2021-10-21]
32. Watson B, Robinson DHZ, Harker L, Arriola KRJ. The inclusion of African American study participants in web-based research studies: Viewpoint. J Med Internet Res 2016 Jun;18(6):e168 [FREE Full text] [doi: 10.2196/jmir.5486] [Medline: 27334683]
33. Lohse B. Facebook is an effective strategy to recruit low-income women to online nutrition education. J Nutr Educ Behav 2013 Jan;45(1):69-76. [doi: 10.1016/j.jneb.2012.06.006] [Medline: 23305805]
34. Why was I selected? US Centers for Disease Control and Prevention. 2021. URL: https://www.cdc.gov/nchs/nhanes/participant/participant-selected.htm [accessed 2021-07-12]
35. Pablos-Méndez A, Barr RG, Shea S. Run-in periods in randomized trials: Implications for the application of results in clinical practice. JAMA 1998 Jan 21;279(3):222-225. [doi: 10.1001/jama.279.3.222] [Medline: 9438743]
36. Dickert N, Grady C. What's the price of a research subject? Approaches to payment for research participation. N Engl J Med 1999 Jul 15;341(3):198-203. [doi: 10.1056/NEJM199907153410312] [Medline: 10403861]
37. Provisional COVID-19 deaths by race and Hispanic origin, and age. US Centers for Disease Control and Prevention. URL: https://data.cdc.gov/w/ks3g-spdg/tdwk-ruhb?cur=i5ViW9i5uLW&from=iTBz4AXWaWh [accessed 2021-03-23]
38. QuickFacts: United States. United States Census Bureau. URL: https://www.census.gov/quickfacts/fact/table/US/RHI125219 [accessed 2020-07-15]
39. Community Engaged Research Initiative. Duke Clinical & Translational Science Institute. URL: https://ctsi.duke.edu/community/community-engaged-research-initiative [accessed 2021-02-25]
40. Baca-Motes K, Edwards AM, Waalen J, Edmonds S, Mehta RR, Ariniello L, et al. Digital recruitment and enrollment in a remote nationwide trial of screening for undiagnosed atrial fibrillation: Lessons from the randomized, controlled mSToPS trial. Contemp Clin Trials Commun 2019 Jun;14:100318 [FREE Full text] [doi: 10.1016/j.conctc.2019.100318] [Medline: 30656241]
41. Liu Y, Pencheon E, Hunter RM, Moncrieff J, Freemantle N. Recruitment and retention strategies in mental health trials - A systematic review. PLoS One 2018;13(8):e0203127 [FREE Full text] [doi: 10.1371/journal.pone.0203127] [Medline: 30157250]
42. Topolovec-Vranic J, Natarajan K. The use of social media in recruitment for medical research studies: A scoping review. J Med Internet Res 2016 Nov 07;18(11):e286 [FREE Full text] [doi: 10.2196/jmir.5698] [Medline: 27821383]
43. Treweek S, Pitkethly M, Cook J, Fraser C, Mitchell E, Sullivan F, et al. Strategies to improve recruitment to randomised trials. Cochrane Database Syst Rev 2018 Dec 22;2:MR000013. [doi: 10.1002/14651858.MR000013.pub6] [Medline: 29468635]

Abbreviations

BYOD: bring your own device
CTSI: Clinical and Translational Science Institute
mHealth: mobile health
NIH: National Institutes of Health

Edited by L Buis; submitted 23.04.21; peer-reviewed by N De Witte, J Claggett; comments to author 04.06.21; revised version received 30.07.21; accepted 15.12.21; published 08.04.22.

Please cite as:
Cho PJ, Yi J, Ho E, Shandhi MMH, Dinh Y, Patil A, Martin L, Singh G, Bent B, Ginsburg G, Smuck M, Woods C, Shaw R, Dunn J
Demographic Imbalances Resulting From the Bring-Your-Own-Device Study Design
JMIR Mhealth Uhealth 2022;10(4):e29510
URL: https://mhealth.jmir.org/2022/4/e29510
doi:10.2196/29510
PMID:34913871

Viewpoint

Interaction Empowerment in Mobile Health: Concepts, Challenges, and Perspectives

Marietta Hamberger[1*], MSc; Nensi Ikonomi[1*], MSc; Julian D Schwab[1*], PhD; Silke D Werle[1*], PhD; Axel Fürstberger[1], PhD; Angelika MR Kestler[2], MD; Martin Holderried[3], MD, PhD; Udo X Kaisers[4], MD; Florian Steger[5], MD; Hans A Kestler[1], PhD

[1]Institute of Medical Systems Biology, Ulm University, Ulm, Germany

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

[3]Department of Medical Development and Quality Management, University Hospital Tübingen, Tübingen, Germany

[4]Chief Executive Officer, University Hospital Ulm, Ulm, Germany

[5]Institute of the History, Philosophy and Ethics of Medicine, Ulm University, Ulm, Germany

[*]these authors contributed equally

Corresponding Author:
Hans A Kestler, PhD
Institute of Medical Systems Biology
Ulm University
Albert-Einstein-Allee 11
Ulm, 89081
Germany
Phone: 49 731 500 ext 24500
Fax: 49 731 50024502
Email: hans.kestler@uni-ulm.de

Abstract

In its most trending interpretation, empowerment in health care is implemented as a patient-centered approach. In the same sense, many mobile health (mHealth) apps are being developed with a primary focus on the individual user. The integration of mHealth apps into the health care system has the potential to counteract existing challenges, including incomplete or nonstandardized medical data and lack of communication, especially in the intersectional context (eg, patients, medical forces). However, concerns about data security and privacy, regional differences in regulations, lack of accessibility, and nontransparent apps hinder the successful integration of mHealth into the health care system. One approach to address this is to rethink the interpretation of empowerment. On that basis, we here examine existing approaches of individual empowerment and subsequently analyze a different view of empowerment in digital health, namely interaction empowerment. Such a change of perspective could positively influence intersectoral communication and facilitate secure data and knowledge sharing. We discuss this novel viewpoint on empowerment, focusing on more efficient integration and development of mHealth approaches. A renewed interpretation of empowerment could thus buffer current limitations of individual empowerment while also advancing digitization of the health system.

(JMIR Mhealth Uhealth 2022;10(4):e32696) doi:10.2196/32696

KEYWORDS

mHealth; mobile apps; patient empowerment; digital health; interaction empowerment; patient-doctor relationship; health care network; intersectoral communication

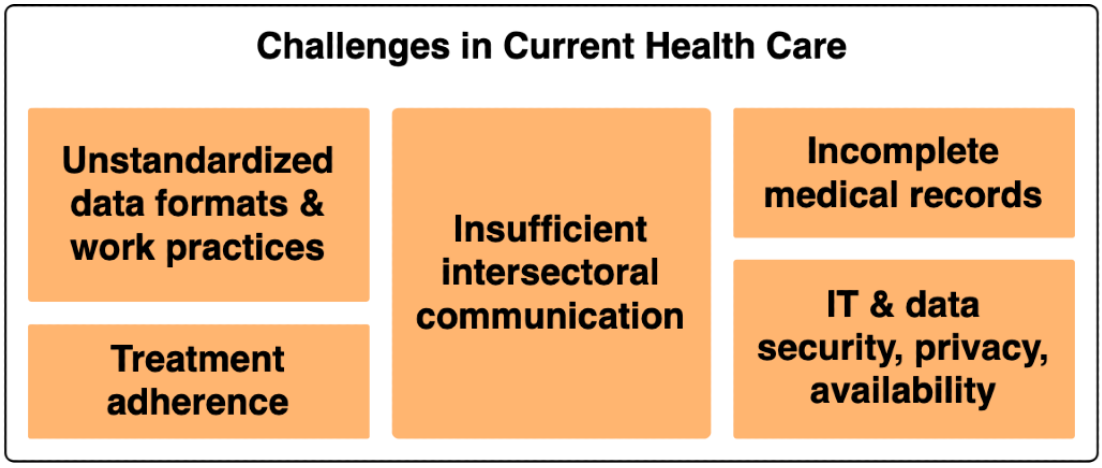
Background

Mobile health (mHealth) represents reformed health and medical care achieved through mobile devices. In the clinical setting or outside hospital walls [1], mHealth is meant to improve and support the prevention, diagnostics, therapy, monitoring, and follow-up care of and with patients [2-5]. Infomediaries, mobile apps, telemedicine, and mobilized medical devices are examples of how mHealth has already been introduced into medical care [3,6-10]. Despite these benefits, limitations due to security concerns, possible health risks, corresponding regulations, and individual barriers influence the development of mHealth and its integration into existing clinical structures [11-20].

Within current clinical settings, common challenges are incomplete medical records [21], data security and privacy [22], as well as unstandardized data formats or work practices [16,23]. Furthermore, missing treatment adherence [24-26] and insufficient communication among physicians, patients, and other health care providers (HCPs) may impede medical care [27-29], as shown in Figure 1.

To address these challenges, we here provide a new viewpoint on the inclusion of empowerment concepts when developing mHealth apps. By summarizing current concepts and challenges in digital medicine, we provide solutions and examples on how to ensure an interaction network–based view on empowerment. Our viewpoint attempts to encourage new discussion and perspectives in implementation of mHealth apps within the current complex health care network of involved parties.

Figure 1. Major challenges in the current health care system. IT: information technology.



Context and Challenges in mHealth Apps

mHealth inter alia aims at improving the challenges of current health care. Through mHealth, various health conditions such as chronic diseases [30], acute conditions [1], mental health disorders [10], or nonillness-related motives are addressed [6]. Many of the current approaches mainly focus on empowerment of the individual person, especially on patient empowerment. For instance, appointment managers [9] and infomediaries already exist for enhanced (educational) communication [8]. Moreover, patient-managed symptom and medication diaries [7,10] or implemented notifications regarding important events are used to improve treatment adherence or counter incomplete medical records [10]. Although the responsibility for treatment decisions for children remains with parents or guardians, this integration process aims to accomplish informed consent by involving minors in the decision-making process [7,30].

Nonetheless, mHealth approaches also entail further barriers that can roughly be separated into personal, technical, and environmental and organizational barriers, as concluded by Schreiweis et al [31]. In Figure 2, we extend this categorization by incorporating additional hurdles collected from the literature [11-20]. While some barriers might hinder the individual user from being empowered, technical or environmental and organizational barriers might influence the necessary interactions among those individuals. As an example, missing standards might obstruct interaction among physicians, developers, and researchers. On the other side, mHealth could also simplify the introduction of standardized practices in health care. However, the actual use would depend on the willingness of individual organizations or people.

Moreover, the sheer number of existing standards and their regional differences may further compromise this approach [32]. The internal competition between the quality of the service and the financial expenses might become disadvantageous for the patient, which is especially relevant in countries dominated by privatized health care services (eg, the United States) [33,34]. Given the rising costs of digitalization processes in health care services, mHealth apps can become a key point on future reduction of financial efforts. In this direction, open-source apps can help in reducing the overall expenses for individual institutions [35].

Although numerous mHealth apps are now freely available for download in leading app stores, only a handful have been designed in accordance with current regulations or guidelines or have been approved by an ethics committee [36,37]. Aside from significant security concerns, this may lead to discouragement among users, who may not be able to differentiate which services best suit their needs [36].

Another obstacle to the global health system benefiting from digitalization is the scarcity of evidence reporting on mHealth interventions [38,39]. Adding to this difficulty, the few existing reports display a broad variety of quality, detail, and objectivity.

To enable accurate evaluation and reproducibility, standardized reporting should also be considered when developing and testing mHealth strategies. The World Health Organization (WHO) mHealth Technical Evidence Review Group devised a checklist for reporting and assessing mHealth evidence. In this checklist, the intervention's scope and context, as well as its technical foundation, may be outlined in a predefined structure [37].

In addition to costs and infrastructure organization issues, most of the challenges in integrating health care and mHealth approaches rely on the involvement of stakeholders. A classical view of the mHealth development and implementation process sees patients as the primary users and, therefore, the empowerment focus. However, in a more realistic scenario, all

of the components in the surrounding health care system should also be considered [40-42].

The network underlying the current health care system comprises different roles. As depicted in Figure 3, a representation of this network can be a complex graph with many participants and interactions. Many representations of the health care system depict the patient in the center or do not allow for an iterative information flow due to a one-directional or tree-like architecture [43,44]. Despite the fact that the patient's well-being is of paramount significance, we wish to emphasize the importance of *all* actors and their interactions via this decentralized interpretation. Toward this end, we adopted a circular design in an attempt to display the relevance of each participant in a more balanced manner.

In the following, the term "node" is used to refer to a specific group of individuals, such as physicians, while "edges" describe the node's interaction (eg, the patient-physician relationship). The edges in Figure 3 represent the direct interactions that we evaluated as relevant, meaning that we excluded, for example, the interaction between relatives and developers due to the uncommonness. Nevertheless, the interaction between patient and relatives may play a role in health care since relatives may be able to observe significant events or provide emotional support. Here, and in the following, we consider two types of relatives. On the one hand, we refer to close contacts, as they are involved in the supportive care of patients among other roles. On the other hand, we consider legal guardians of a patient if they have the final decision-making power, as in the case of children. To simplify this separation, we use the term "relatives" to refer to the first group and the term "patient" to denote the individuals making the decision (ie, a legal guardian), if applicable.

Figure 2. Three main classes of barriers in designing mobile health apps as categorized by Schreiweis et al [31]. For each larger class, subboxes including specific issues are depicted. Boxes with solid borders represent barriers mentioned by Schreiweis et al [31]. Additional or more detailed hurdles we collected from the literature are highlighted by dashed lines [11-20]. The size of the boxes does not represent their relevance or severity. AI: artificial intelligence.

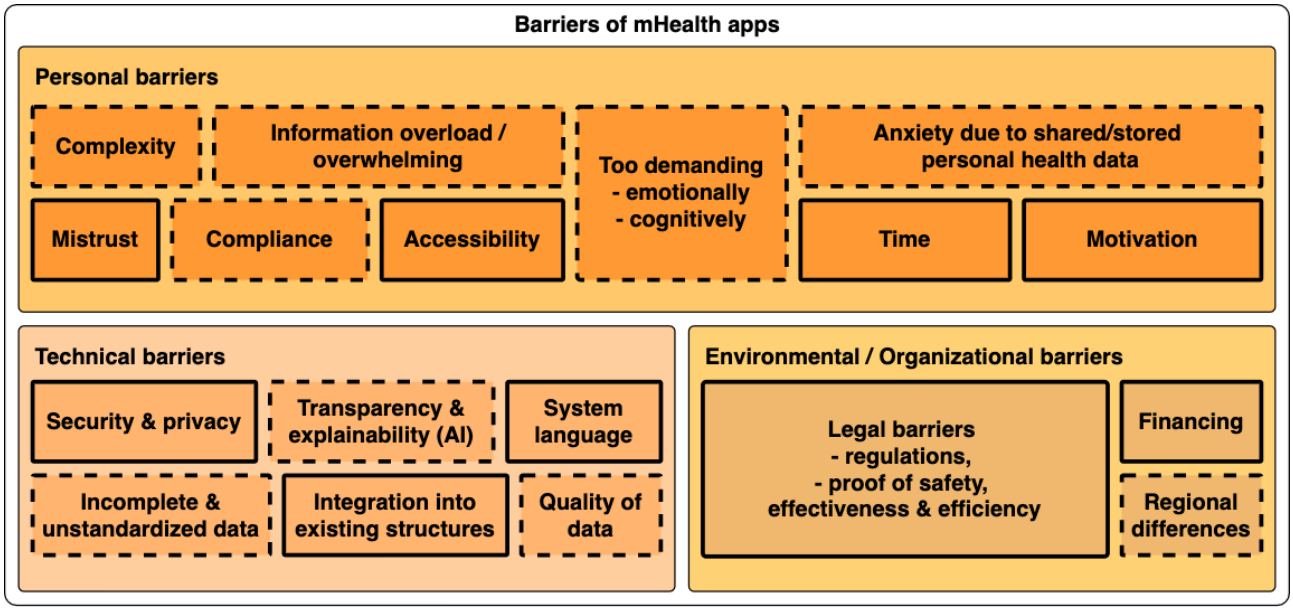
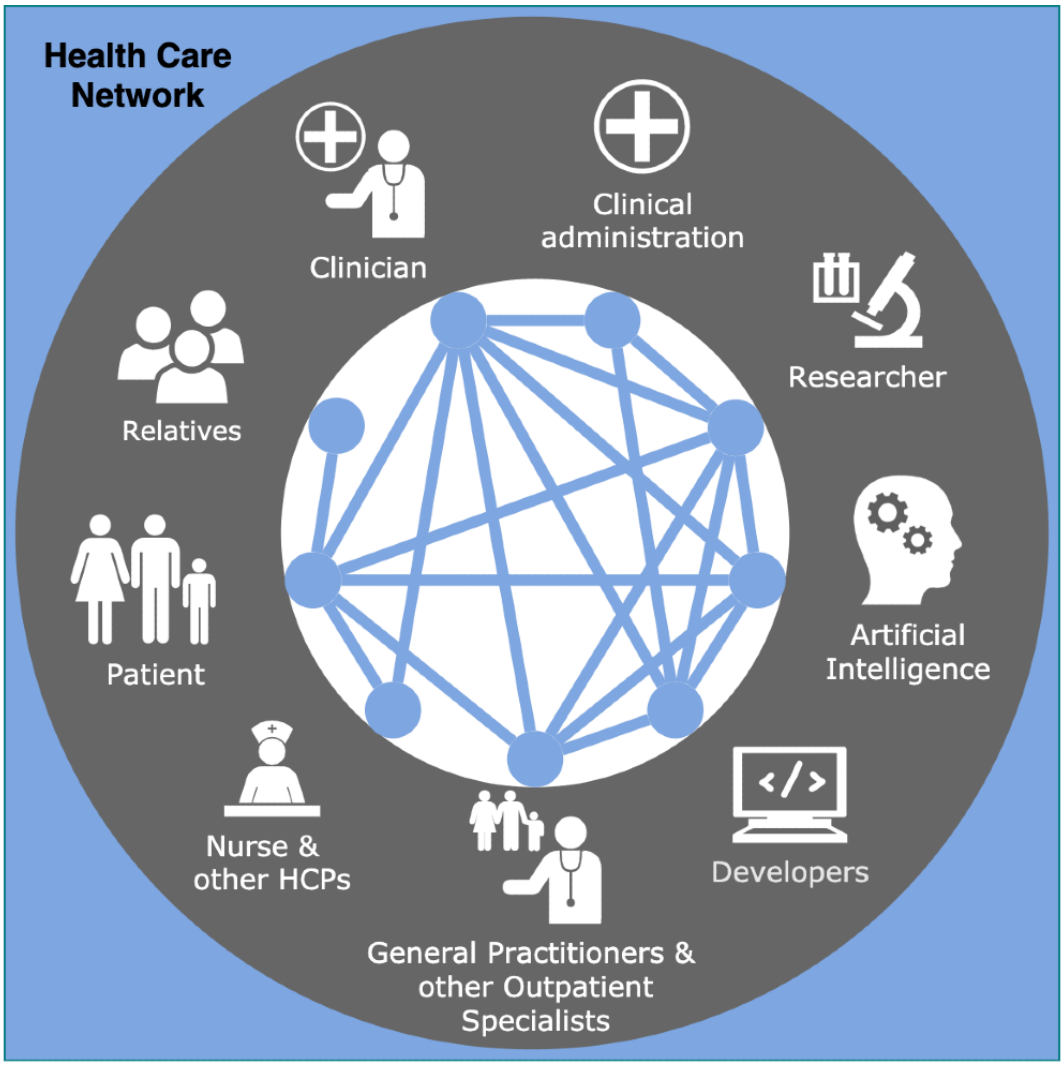


Figure 3. Representation of the modern health care network. Each node represents a relevant group of individuals interacting within this network with other individuals. Connecting edges represent these possible interactions. Even though there are more actors in the existing health care system, we merely included the groups we evaluated to be the most relevant for the concept of empowerment. HCP: health care provider.



Empowerment of Individual Nodes in the Health Care Network

Overview

Ideally, the patient's health should be seen as the main objective of the health care network. Nonetheless, each actor may follow their own personal objective while trying to gain power and fearing losing it while providing or receiving health care.

The Patient

We assume that the patients' primary goals are to maintain or improve their health, manage their lives, and receive adequate support after receiving a diagnosis. A hierarchical or linear organization of health care (ie, putting the more medically educated HCPs at the beginning or top and the patient at the receiving end) might be outdated. Recently, patient-centered care has been considered to be more beneficial [45]. Given the trend of patient empowerment, it is natural to picture the empowered patient in the center with supporting individuals and HCPs acting in the surroundings. Regarding shared decision-making, most patients prefer to be included in the decision-making process. Nonetheless, approximately half of all patients would prefer that the physician should have the final say, even though this case is legally unenforceable since physicians are merely acting on behalf of the patient. Therefore, the average patient strives to be included in the decision-making process, yet the preferences regarding the final decision would theoretically vary [46]. Furthermore, the patients' preference for playing an active role in the treatment process depends on different factors, including their age [46], ethnicity, or level of education [47]. Some patients do not want to be confronted by their health condition or do not feel that they are sufficiently ill to take action, leading to a potential delay in reaching out to HCPs [48-50]. Hence, patients may differ in their preferences concerning decision-making or being educated. Nevertheless, the patient, or their legal guardian, is obligated to decide, which emphasizes the importance of sufficient education and information.

Artificial Intelligence

In Figure 3, we also included artificial intelligence (AI) as a relevant interacting node. Recently gaining popularity, AI is already part of various health-related apps and devices [51]. AI may also be integrated into mHealth, yet current algorithms are

still too prone to errors. Hence, at this point in development, they cannot be trusted with medical decisions (eg, due to not being able to interpret social cues or engaging in personal discussions with the respective patient) [52]. Nonetheless, the continuous improvement of such apps justifies the use of AI in an assisting manner. In this context, the transparency and explainability of these systems is crucial to ensure human users' trust in AI as assistants [52,53]. An explainable app assists the users and reveals why it made a specific assessment. A user with domain knowledge can then interfere and perform error correction in case of erroneous results. This approach may also help to detect uncommon lesions, which might be overseen due to the inexperience of the individual clinician.

Moreover, the AI screening approach may save time for the clinician, which can be redirected to patient consultation and more direct contact. Using a conversational interface, chatbots can, for example, provide an initial evaluation of a patient's mental or physical state or empower individuals confronted with a life-altering condition [54,55]. Besides alleviating the frequent concern of patients to lose personal contact due to digital medical approaches [56], such approaches linking the AI to the clinician can further boost the doctor-patient relationship. Moreover, while the user benefits from an AI assistant without giving up control, the AI benefits from the chance of improvement through error correction [53]. In this sense, this interaction can be seen as mutually empowering if accessibility is put into practice to counterbalance social inequalities. Yet, it is essential to note that additional challenges accompany the benefits of AI integration. For this purpose, the WHO guidance on Ethics & Governance of Artificial Intelligence for Health promotes an ethical integration of AI into digital health care solutions based on six core principles [57], including discussing the transparency of AI in health care. Additionally, biased databases are problematic since they are often based primarily on male subjects and usually do not include minorities, possibly limiting the validity of the data [58,59].

Health Care Providers

Physicians, either practicing in primary care or a clinical environment, must still focus on compassionate healing as the foundation of health care, even though the mode of interacting may change. Frequently, continuity of care is impeded by insufficient data sharing among physicians and other HCPs (eg, information about treatment decisions). Simultaneously, being equipped with more time, opportunity, information, and technologies, the training of the empowered physician to treat and monitor remotely is also essential [17]. Furthermore, the voice of nursing and the viewpoint of frontline workers is described as a critical point in developing mHealth apps [18]. In this context, the empowerment of nurses and community health workers (CHWs) can positively transfer to their patients [9,18].

Problems of Empowering the Individual in mHealth

Even though the traditional doctor-patient relationship may leave the patient with less control and provided knowledge, patient-centered health care combined with the hard-to-balance empowerment of individual roles can result in a fight-or-flight reaction among physicians. While in this movement, the patient-physician relationship changes into a partnership, with the physician's role slowly shifting into guiding the patient in shared decision-making [17], physicians might see their clinical autonomy threatened. However, the patient's trust in digital solutions may be influenced by the physician's opinion and recommendation [11,60]. In addition, personal factors can influence mHealth app usage, such as age, general preferences, or educational background [61]. The resulting physicians' disempowerment can thus counter the anticipated advancements. Thus, the empowerment of an individual node may positively influence the empowerment of another in the health care network. Nonetheless, the exact opposite now poses a parallel challenge, where empowerment of one individual triggers disempowerment of the other. Additionally, the individuals' empowerment through mHealth does not directly tackle the challenges of current health care mentioned above (see Figure 1).

Even though individual empowerment might positively influence treatment adherence, the other challenges remain primarily unaffected. The reason for this may rely on the fact that these other problems emerge when individuals are interacting (eg, while communicating, through data exchange, or teamwork). Since the approach of empowering the individual merely focuses on the network's nodes, the edges in between nodes are not prioritized. Thus, instead of empowering single nodes of the health care network, we recommend strengthening the interactions, thereby empowering the edges. This change of viewpoint may enhance the ability to develop approaches tackling the interaction-related challenges.

Empowerment of the Interaction Network

Based on the issues mentioned above, we propose a new view on empowerment, namely the *empowerment of interactions*. All interactions follow a specific goal that is mostly shared among all participants, such as the improvement of a patient's acute health status or the development of an app [62].

Through a pooled skill set and empowered mutual support, the underlying goal might be reached more efficiently. Instead of trying to improve each individual's skill set, the weaknesses of one party could be directly balanced out by the strengths and skills of another [63]. Hence, technical skills, medical knowledge, information about patient history, administrative and legal regulations, and standard practices can be actively shared at the point of need, in contrast to time-consuming preeducation. Here, in empowering the interaction, the redistribution of power and control, as well as the reshaping of single relationships, is not essential yet possible. As an example, privacy and security issues have to be considered within their

network of influence. For instance, by empowering communication between developers and HCPs in defining safety priorities, patients will also benefit in trust when using the mobile app. Hence, instead of a technical implementation, data security becomes a key factor within empowering trust and safety of involved parties [64].

Thus, this approach can facilitate integration into traditional structures and offers the benefits of individual empowerment, such as better-informed patients or improved continuity of care.

Until this point, we have considered a general network of multiple individuals with specific roles that interact with others in various ways, all depending on the interaction's goal. Nevertheless, the concept of interaction empowerment is applicable on multiple levels. For example, the simplest way would be to empower the relationship between two individuals. Yet, on a more abstract and broad level, one could empower multiple interactions simultaneously. In pooling single interactions together to form subnetworks following a common goal, we can again focus on one entity for empowerment, namely the subnetwork.

Interactions in the Health Care Network

Defining Subnetworks

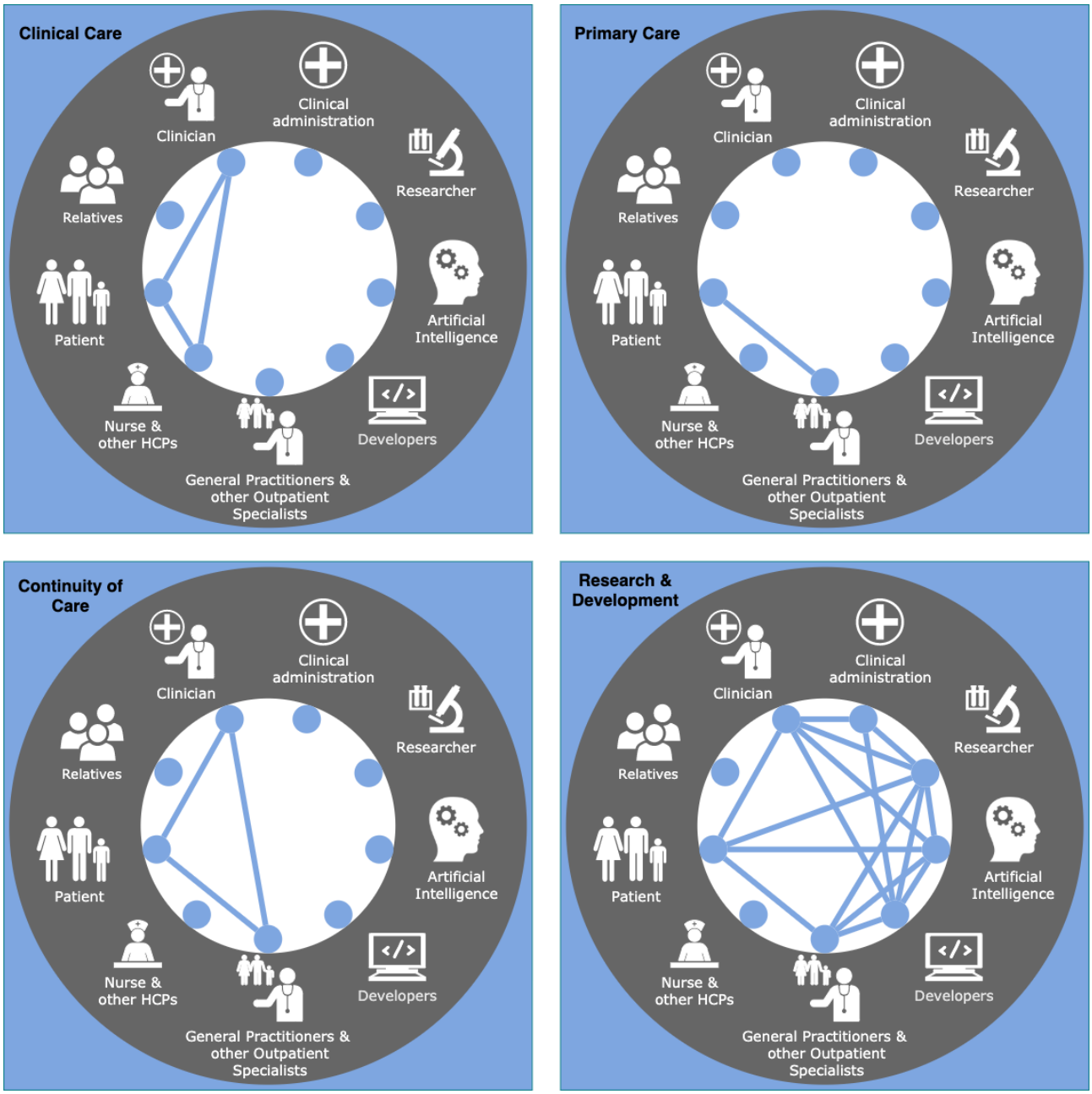
Various interactions take place in the health care network (see Figure 3). Within the network, individuals communicate, share data and knowledge, cooperate, or provide treatment and support [65,66]. While interacting, each individual has a specific role comprising their own skill sets, motives, and more personal characteristics. Additionally, the same individuals may interact in different ways based on their goal of interaction. Hence, the individuals' roles and interactions may change depending on the underlying interaction type and intention (eg, primary care or research and development) [67,68]. Exemplarily, a physician might be treating a patient (primary care), whereas in another context they can contribute to research projects (development). While the primary aim relies on empowering the network of stakeholders, single nodes still retain decision power on downregulating or eliminating a specific interaction. This implies that individual nodes (stakeholders) and edges (empowerment means) can crosstalk in defining the final empowerment network.

Empowering the interaction does not merely mean reshaping or prioritizing relationships. Instead, such empowerment can be defined as identifying the therein present interactions, their underlying goals, and relevant actors, and determining how these interactions can be supported in a balanced manner. Given the multilayers and roles described above, the overall interaction network could end up being a complex interplay of intricate edges. Hence, we propose to approach interaction networks merely as a channel for information flow, independent of the types and number of participants. This allows for the formation of generalized constructs, which can be resolved in generating functional subnetworks for each empowerment task. For this purpose, we next aim at partitioning the health care network into functional subnetworks.

Putting these concepts into practice, in principle, almost all nodes of the health care network may interact with each other for multiple contexts and goals (see Figure 3). Following the idea of the communication flow, subnetworks can be identified based on specific tasks (see Figure 4). Subnetworks can overlap in part of their nodes and interactions, still yielding a different empowering task. In addition, one subnetwork may be entirely part of another subnetwork regarding their nodes and edges, yet aiming at another or more specific goal. Exemplarily, the primary care subnetwork and the continuity of care subnetwork share common nodes and edges, even if the final empowerment task is set to different aims.

Through pooling instances (ie, nodes and edges following the same goal), their weaknesses may be balanced out by the strengths and skills of other participating individuals. With this approach, the mentioned benefits of interaction empowerment are still incorporated, yet the count of empowerment targets is reduced to the count of chosen subnetworks. In summary, we treat the subnetwork as one instance following one goal with a set of strengths and weaknesses that may cancel each other out to a certain degree.

Figure 4. Four examples of possible subnetworks in the modern health care network. Each node represents a relevant role (ie, individual or group) and the connecting edges represent relevant interactions in the underlying subnetwork. The top left caption describes the interaction's main aim. HCP: health care provider.



Use Cases of Subnetworks in Health Care

To provide practical use cases, we describe four chosen subnetworks, as depicted in Figure 4: clinical care, primary care, continuity of care, and research and development. To each subnetwork, we map the challenges of current health care that we assume to be positively influenced by it.

In the first subnetwork, patients, clinicians, nurses, and other HCPs interact for clinical health care. Here, lacking treatment adherence might obstruct successful clinical care. The empowerment of this network might help to tackle this challenge.

Similarly, the patient may interact with the family physician for primary care (ie, for day-to-day health care). Here, the lack of treatment adherence also poses a challenge. In this specific case, the network generally only comprises two nodes with one interaction entity. We nonetheless consider this interaction as relevant to form an individual subnetwork.

Third, the interaction among clinicians, the family physician, and the patient ensures continuity of care, as in the context of a hospital stay. However, unstandardized data formats and work practices, data security, insufficient intersectoral communication, and resulting incomplete medical records hinder the continuity of care. These challenges may, in our opinion, be tackled by empowering this subnetwork.

Finally, in research and development, researchers, developers, physicians (clinicians/family physicians), patients, and clinical administration may interact, with each node's contribution and degree of involvement depending on the unique context/setting.

Here, the researcher, developer, or clinician might present as the same individual, such as in the field of personalized medicine. Academia-industry cooperation and institutional review boards can also fall into this fourth category. Herein, the progress is slowed down by data security and privacy issues and insufficient intersectoral communication. Again, by empowering this subnetwork's interactions, we argue that those challenges may be improved.

Empowering Subnetworks Through mHealth

Some mHealth apps already support a concept of empower-oriented communication through interactions. In the following, we summarize some examples tackling different network empowerment concepts, which can be used as a starting point for further development.

CoCoV [69] is an mHealth app that supports patients to track their symptoms after receiving a COVID-19 vaccination through an easily understandable questionnaire. This app empowers the physician-patient interaction in preventing lost or overlooked symptoms. Additionally, the safely and anonymously transferred structured data can further be collected, enabling enhanced analysis of adverse symptoms by researchers and data scientists. Hence, the design of CoCoV simultaneously tackles two different empowerment subnetworks.

The app NEMO [69,70] supports patients receiving tumor therapy to protocol-related side effects in a structural and standardized manner. Typically, patients may forget to record incidents or may not consider them relevant [71]. The management app hence empowers the patient-physician interaction and communication in counteracting those issues. Since patients can administer the data sharing through QR codes, they can be empowered individually and continuity of care can be enhanced. NEMO [69,70] can be further used to empower the treating physician and other HPCs by facilitating data sharing (eg, among family physicians and specialists). Additionally, as with CoCoV, standardized and anonymized data collection empowers research and development.

Track Your Tinnitus is another symptom tracker enabling the patient to record daily tinnitus perception [72]; the app PainBuddy focuses on pain and symptom management of children with cancer [7]; and MindFrame empowers patients recently diagnosed with schizophrenia in managing symptoms, medication, and recurring behaviors [10]. Like CoCoV and NEMO, these apps empower the individual yet also support the interaction, communication, and data transfer, resulting in interaction empowerment of the respective subnetworks. Focusing more on empowered communication, Ramukumba et al [9] presented an mHealth app used by CHWs. Here, the support for communication through information exchange and timely decision support at the point of care resulted in the empowerment of CHWs, and consequently that of their patients.

FeelBack is a project assessing psycho-oncological stress through an app-based questionnaire. Here, not only patients can be targeted but also their relatives feeling burdened by the illness. Hence, by empowering the related individuals, the supportive interaction among patients and others can be empowered [19].

In simplifying intersectoral communication and data sharing, the browser-based system CTest enables tracking clinical samples online [73]. The tested individual can access their test sample's status update through personalized QR codes. With this approach, clinical staff can be disburdened while the testee is not dependent on clinical staff for accessing test results. Hence, this interaction is empowered through simplification of the overall communication process.

Conclusion

Recently, mHealth apps are being developed to aid with specific tasks in the health care system. The empowerment of the individual (ie, patient or physician) is often seen as a central aspect. Nonetheless, this approach entails further challenges and barriers such as disempowerment through empowerment, difficulties in integrating mHealth into existing structures, or social inequalities. We have presented an updated approach to instead focus on the *empowerment of interaction and communication*, since we argue that this could help to tackle current challenges in the health care system even faster and more efficiently. To reach a specific goal (eg, continuity of care), multiple groups in the health care network interact, which can be pooled to a single subnetwork. In developing mHealth with a strengthened focus on empowerment of interaction networks, the individual may still benefit even if the focus is shifted to another level, while also diminishing the barriers of the original individualized approach. Hence, we propose a new perspective on mHealth and empowerment to tackle current health care challenges more efficiently.

Acknowledgments

HK acknowledges funding provided by the Ministry of Science and Art Baden-Württemberg (Zentrum für Innovative Versorgung); the German Federal Ministry of Education and Research, as part of the DIFUTURE project (Medical Informatics Initiative, grant 01ZZ1804I); and the Ministry of Social Affairs and Integration Baden-Württemberg as part of the project feelBack (networked, digital, patient-related psycho-oncology feedback).

Authors' Contributions

HK is responsible for funding acquisition and was responsible for supervision of the project. MH, NI, JS, SW, and HK were responsible for visualization of the results. MH, NI, JS, SW, and HK were responsible for writing the original draft of the manuscript. All authors were responsible for writing, discussing, reviewing, and editing the manuscript.

Conflicts of Interest

None declared.

References

1. Nix J, Comans T. Home Quick - occupational therapy home visits using mHealth, to facilitate discharge from acute admission back to the community. Int J Telerehabil 2017;9(1):47-54 [FREE Full text] [doi: 10.5195/ijt.2017.6218] [Medline: 28814994]
2. Asadzadeh A, Kalankesh LR. A scope of mobile health solutions in COVID-19 pandemics. Inform Med Unlocked 2021;23:100558 [FREE Full text] [doi: 10.1016/j.imu.2021.100558] [Medline: 33842688]
3. Holderried M, Höper A, Holderried F. Innovationen und Innovationsmanagement im Gesundheitswesen. In: Pfannstiel M, Kassel K, Rasche C, editors. Disruption E-Health: Treiber für die sektorübergreifend-personalisierte Medizin der Zukunft. Wiesbaden: Springer Gabler; 2020:393-415.
4. Osborn J, Ajakaiye A, Cooksley T, Subbe CP. Do mHealth applications improve clinical outcomes of patients with cancer? A critical appraisal of the peer-reviewed literature. Support Care Cancer 2020 Mar;28(3):1469-1479 [FREE Full text] [doi: 10.1007/s00520-019-04945-4] [Medline: 31273501]
5. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med 2020;3:4. [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]
6. Delgado GV, Carvalho RB, Choo CW, Leite RS, Castro JM. Patient empowerment through mobile health: Case study with a Brazilian application for pregnancy support. Proc Assoc Inf Sci Technol 2020 Oct 22;57(1):e221. [doi: 10.1002/pra2.221]
7. Fortier MA, Chung WW, Martinez A, Gago-Masague S, Sender L. Pain buddy: A novel use of m-health in the management of children's cancer pain. Comput Biol Med 2016 Sep 01;76:202-214 [FREE Full text] [doi: 10.1016/j.compbiomed.2016.07.012] [Medline: 27479493]
8. Khuntia J, Yim D, Tanniru M, Lim S. Patient empowerment and engagement with a health infomediary. Health Policy Technol 2017 Mar;6(1):40-50. [doi: 10.1016/j.hlpt.2016.11.003]
9. Ramukumba MM, Hägglund M. "I feel like a nurse and my clients learn more": mHealth, capacity building and empowerment in community based care. Stud Health Technol Inform 2019 Aug 09;265:195-200. [doi: 10.3233/SHTI190163] [Medline: 31431598]
10. Terp M, Jørgensen R, Laursen BS, Mainz J, Bjørnes CD. A smartphone app to foster power in the everyday management of living with schizophrenia: qualitative analysis of young adults' perspectives. JMIR Ment Health 2018 Oct 01;5(4):e10157 [FREE Full text] [doi: 10.2196/10157] [Medline: 30274966]
11. Affinito L, Fontanella A, Montano N, Brucato A. How physicians can empower patients with digital tools. J Public Health 2020 Jul 23;30(4):897-909. [doi: 10.1007/s10389-020-01370-4]
12. Fadda M, Galimberti E, Fiordelli M, Romanò L, Zanetti A, Schulz PJ. Effectiveness of a smartphone app to increase parents' knowledge and empowerment in the MMR vaccination decision: A randomized controlled trial. Hum Vaccin Immunother 2017 Nov 02;13(11):2512-2521 [FREE Full text] [doi: 10.1080/21645515.2017.1360456] [Medline: 29125783]
13. Guo C, Ashrafian H, Ghafur S, Fontana G, Gardner C, Prime M. Challenges for the evaluation of digital health solutions-A call for innovative evidence generation approaches. NPJ Digit Med 2020;3(1):110. [doi: 10.1038/s41746-020-00314-2] [Medline: 32904379]
14. Jungmann SM, Brand S, Kolb J, Witthöft M. Do Dr. Google and health apps have (comparable) side effects? An experimental study. Clin Psychol Sci 2020 Mar 04;8(2):306-317. [doi: 10.1177/2167702619894904]
15. Kondylakis H, Bucur A, Crico C, Dong F, Graf N, Hoffman S, et al. Patient empowerment for cancer patients through a novel ICT infrastructure. J Biomed Inform 2020 Jan;101:103342 [FREE Full text] [doi: 10.1016/j.jbi.2019.103342] [Medline: 31816400]
16. Lehne M, Sass J, Essenwanger A, Schepers J, Thun S. Why digital medicine depends on interoperability. NPJ Digit Med 2019;2:79. [doi: 10.1038/s41746-019-0158-1] [Medline: 31453374]
17. Mesko B, Győrffy Z. The rise of the empowered physician in the digital health era: viewpoint. J Med Internet Res 2019 Mar 26;21(3):e12490 [FREE Full text] [doi: 10.2196/12490] [Medline: 30912758]
18. Samples C, Ni Z, Shaw RJ. Nursing and mHealth. Int J Nurs Sci 2014 Dec;1(4):330-333. [doi: 10.1016/j.ijnss.2014.08.002]
19. Schobel J, Volz M, Hörner K, Kuhn P, Jobst F, Schwab JD, et al. Supporting medical staff from psycho-oncology with smart mobile devices: insights into the development process and first results. Int J Environ Res Public Health 2021 May 11;18(10):5092 [FREE Full text] [doi: 10.3390/ijerph18105092] [Medline: 34064987]
20. Wangler J, Jansky M. The use of health apps in primary care-results from a survey amongst general practitioners in Germany. Wien Med Wochenschr 2021 May;171(7-8):148-156 [FREE Full text] [doi: 10.1007/s10354-021-00814-0] [Medline: 33570692]
21. Cook RI, Render M, Woods DD. Gaps in the continuity of care and progress on patient safety. BMJ 2000 Mar 18;320(7237):791-794 [FREE Full text] [doi: 10.1136/bmj.320.7237.791] [Medline: 10720370]
22. Knoppers BM, Thorogood AM. Ethics and big data in health. Curr Opin Syst Biol 2017 Aug;4:53-57. [doi: 10.1016/j.coisb.2017.07.001]

23. Gonzales M, Cheung V, Riek L. Designing collaborative healthcare technology for the acute care workflow. 2015 Presented at: PervasiveHealth 2015. 9th International Conference on Pervasive Computing Technologies for Healthcare; May 20–23, 2015; Istanbul, Turkey p. 145-152. [doi: 10.4108/icst.pervasivehealth.2015.259287]
24. Cramer JA. A systematic review of adherence with medications for diabetes. Diabetes Care 2004 May;27(5):1218-1224. [doi: 10.2337/diacare.27.5.1218] [Medline: 15111553]
25. Huser MA, Evans TS, Berger V. Medication adherence trends with statins. Adv Ther 2005;22(2):163-171. [doi: 10.1007/BF02849887] [Medline: 16020406]
26. Senior V, Marteau TM, Weinman J, Genetic Risk Assessment for FH Trial (GRAFT) Study Group. Self-reported adherence to cholesterol-lowering medication in patients with familial hypercholesterolaemia: the role of illness perceptions. Cardiovasc Drugs Ther 2004 Nov;18(6):475-481. [doi: 10.1007/s10557-004-6225-z] [Medline: 15770435]
27. Ha JF, Longnecker N. Doctor-patient communication: a review. Ochsner J 2010;10(1):38-43 [FREE Full text] [Medline: 21603354]
28. Lotfi M, Zamanzadeh V, Valizadeh L, Khajehgoodari M. Assessment of nurse-patient communication and patient satisfaction from nursing care. Nurs Open 2019 Jul;6(3):1189-1196. [doi: 10.1002/nop2.316] [Medline: 31367445]
29. Vermeir P, Vandijck D, Degroote S, Peleman R, Verhaeghe R, Mortier E, et al. Communication in healthcare: a narrative review of the literature and practical recommendations. Int J Clin Pract 2015 Nov;69(11):1257-1267 [FREE Full text] [doi: 10.1111/ijcp.12686] [Medline: 26147310]
30. Lalloo C, Hundert A, Harris L, Pham Q, Campbell F, Chorney J, et al. Capturing daily disease experiences of adolescents with chronic pain: mHealth-mediated symptom tracking. JMIR Mhealth Uhealth 2019 Jan 17;7(1):e11838 [FREE Full text] [doi: 10.2196/11838] [Medline: 30664472]
31. Schreiweis B, Pobiruchin M, Strotbaum V, Suleder J, Wiesner M, Bergh B. Barriers and facilitators to the implementation of eHealth services: systematic literature analysis. J Med Internet Res 2019 Nov 22;21(11):e14197 [FREE Full text] [doi: 10.2196/14197] [Medline: 31755869]
32. Jogova M, Shaw J, Jamieson T. The regulatory challenge of mobile health: lessons for Canada. Healthc Policy 2019 Feb;14(3):19-28 [FREE Full text] [doi: 10.12927/hcpol.2019.25795] [Medline: 31017863]
33. Porter ME, Teisberg EO. Redefining competition in health care. Harv Bus Rev 2004 Jun;82(6):64-76, 136. [Medline: 15202288]
34. Porter ME, Teisberg EO. Redefining health care: creating value‐based competition on results. Boston, MA: Harvard Business School Press; 2006.
35. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. PLoS One 2017;12(2):e0170581 [FREE Full text] [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]
36. Holfelder M, Mulansky L, Schlee W, Baumeister H, Schobel J, Greger H, et al. Medical device regulation efforts for mHealth apps during the COVID-19 Pandemic—an experience report of Corona Check and Corona Health. J 2021 Jun 16;4(2):206-222. [doi: 10.3390/j4020017]
37. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016 Mar 17;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]
38. Peiris D, Miranda JJ, Mohr DC. Going beyond killer apps: building a better mHealth evidence base. BMJ Glob Health 2018;3(1):e000676 [FREE Full text] [doi: 10.1136/bmjgh-2017-000676] [Medline: 29527353]
39. Feroz A, Kadir MM, Saleem S. Health systems readiness for adopting mhealth interventions for addressing non-communicable diseases in low- and middle-income countries: a current debate. Glob Health Action 2018;11(1):1496887 [FREE Full text] [doi: 10.1080/16549716.2018.1496887] [Medline: 30040605]
40. Guyatt G, Montori V, Devereaux PJ, Schünemann H, Bhandari M. Patients at the center: in our practice, and in our use of language. ACP J Club 2004;140(1):A11-A12. [Medline: 14711297]
41. Bruce C, Harrison P, Giammattei C, Desai S, Sol JR, Jones S, et al. Evaluating patient-centered mobile health technologies: definitions, methodologies, and outcomes. JMIR Mhealth Uhealth 2020 Nov 11;8(11):e17577 [FREE Full text] [doi: 10.2196/17577] [Medline: 33174846]
42. Gimbel R, Shi L, Williams JE, Dye CJ, Chen L, Crawford P, et al. Enhancing mHealth technology in the patient-centered medical home environment to activate patients with type 2 diabetes: a multisite feasibility study protocol. JMIR Res Protoc 2017 Mar 06;6(3):e38 [FREE Full text] [doi: 10.2196/resprot.6993] [Medline: 28264792]
43. Lamprinos I, Demski H, Mantwill S, Kabak Y, Hildebrand C, Ploessnig M. Modular ICT-based patient empowerment framework for self-management of diabetes: Design perspectives and validation results. Int J Med Inform 2016 Jul;91:31-43. [doi: 10.1016/j.ijmedinf.2016.04.006] [Medline: 27185507]
44. Jung O, Glachs D, Strohmeier F, Mulrenin R, Huisman S, Smith I, et al. Empowering diabetes patients with interventions based on behaviour change techniques. Stud Health Technol Inform 2019;260:154-161. [Medline: 31118332]
45. Rathert C, Wyrwich MD, Boren SA. Patient-centered care and outcomes: a systematic review of the literature. Med Care Res Rev 2013 Aug;70(4):351-379. [doi: 10.1177/1077558712465774] [Medline: 23169897]

46. Levinson W, Kao A, Kuby A, Thisted RA. Not all patients want to participate in decision making. A national study of public preferences. J Gen Intern Med 2005 Jun;20(6):531-535 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.04101.x] [Medline: 15987329]
47. Street RL, Gordon HS, Ward MM, Krupat E, Kravitz RL. Patient participation in medical consultations: why some patients are more involved than others. Med Care 2005 Oct;43(10):960-969. [doi: 10.1097/01.mlr.0000178172.40344.70] [Medline: 16166865]
48. Byrne SK. Healthcare avoidance: a critical review. Holist Nurs Pract 2008;22(5):280-292. [doi: 10.1097/01.HNP.0000334921.31433.c6] [Medline: 18758277]
49. Hardavella G, Aamli-Gaagnat A, Frille A, Saad N, Niculescu A, Powell P. Top tips to deal with challenging situations: doctor-patient interactions. Breathe 2017 Jun;13(2):129-135 [FREE Full text] [doi: 10.1183/20734735.006616] [Medline: 28620434]
50. Taber JM, Leyva B, Persoskie A. Why do people avoid medical care? A qualitative study using national data. J Gen Intern Med 2015 Mar;30(3):290-297 [FREE Full text] [doi: 10.1007/s11606-014-3089-1] [Medline: 25387439]
51. Wang F, Preininger A. AI in health: state of the art, challenges, and future directions. Yearb Med Inform 2019 Aug;28(1):16-26 [FREE Full text] [doi: 10.1055/s-0039-1677908] [Medline: 31419814]
52. Lysaght T, Lim HY, Xafis V, Ngiam KY. AI-assisted decision-making in healthcare: the application of an ethics framework for big data in health and research. Asian Bioeth Rev 2019 Sep;11(3):299-314 [FREE Full text] [doi: 10.1007/s41649-019-00096-0] [Medline: 33717318]
53. Miller T. Explanation in artificial intelligence: insights from the social sciences. Artific Intell 2019 Feb;267:1-38. [doi: 10.1016/j.artint.2018.07.007]
54. Laumer S, Maier C, Gubler F. Chatbot acceptance in healthcare: explaining user adoption of conversational agents for disease diagnosis. 2019 Presented at: 27th European Conference on Information Systems (ECIS); June 8-14, 2019; Stockholm and Uppsala, Sweden. [doi: 10.1145/3322385.3322392]
55. Macedo P, Pereira C, Mota P, Silva D, Frade A, Madeira RN. Conversational agent in mHealth to empower people managing the Parkinson's disease. Procedia Comput Sci 2019;160:402-408. [doi: 10.1016/j.procs.2019.11.074]
56. Holderried M, Schlipf M, Höper A, Meier R, Stöckle U, Kraus TM. Chances and risks of telemedicine in orthopaedics and trauma surgery. Z Orthop Unfall 2018 Feb;156(1):68-77. [doi: 10.1055/s-0043-116941] [Medline: 28837980]
57. World Health Organization. Ethics and governance of artificial intelligence for health. In: WHO guidance. Geneva, Switzerland: WHO; Jun 28, 2021.
58. Holdcroft A. Gender bias in research: how does it affect evidence based medicine? J R Soc Med 2007 Jan;100(1):2-3 [FREE Full text] [doi: 10.1177/014107680710000102] [Medline: 17197669]
59. Niranjan SJ, Martin MY, Fouad MN, Vickers SM, Wenzel JA, Cook ED, et al. Bias and stereotyping among research and clinical professionals: Perspectives on minority recruitment for oncology clinical trials. Cancer 2020 Jan 01;126(9):1958-1968. [doi: 10.1002/cncr.32755] [Medline: 32147815]
60. Holderried M, Hoeper A, Holderried F, Blumenstock G, Ernst C, Tropitzsch A. Attitudes toward e-Health: the otolaryngologists' point of view. Telemed J E Health 2018 Jun;24(6):425-432. [doi: 10.1089/tmj.2017.0158] [Medline: 29227206]
61. Bol N, Helberger N, Weert JCM. Differences in mobile health app use: a source of new digital inequalities? Inf Soc 2018 Apr 26;34(3):183-193. [doi: 10.1080/01972243.2018.1438550] [Medline: 26281194]
62. Marks MA, Mathieu JE, Zaccaro SJ. A temporally based framework and taxonomy of team processes. Acad Manag Rev 2001 Jul;26(3):356-376. [doi: 10.5465/amr.2001.4845785]
63. Siassakos D, Bristowe K, Draycott T, Angouri J, Hambly H, Winter C, et al. Clinical efficiency in a simulated emergency and relationship to team behaviours: a multisite cross-sectional study. BJOG 2011 Apr;118(5):596-607. [doi: 10.1111/j.1471-0528.2010.02843.x] [Medline: 21291509]
64. Iwaya LH, Ahmad A, Babar MA. Security and privacy for mHealth and uHealth systems: a systematic mapping study. IEEE Access 2020;8:150081-150112. [doi: 10.1109/access.2020.3015962]
65. Foronda C, MacWilliams B, McArthur E. Interprofessional communication in healthcare: an integrative review. Nurse Educ Pract 2016 Jul;19:36-40. [doi: 10.1016/j.nepr.2016.04.005] [Medline: 27428690]
66. Pentland D, Forsyth K, Maciver D, Walsh M, Murray R, Irvine L, et al. Key characteristics of knowledge transfer and exchange in healthcare: integrative literature review. J Adv Nurs 2011 Jul;67(7):1408-1425. [doi: 10.1111/j.1365-2648.2011.05631.x] [Medline: 21466576]
67. Schwartz DA. Physician-scientists: the bridge between medicine and science. Am J Respir Crit Care Med 2012 Mar 15;185(6):595-596. [doi: 10.1164/rccm.201110-1806ed]
68. Smith R. Doctors are not scientists. BMJ 2004 Jun 17;328(7454):0-h. [doi: 10.1136/bmj.328.7454.0-h]
69. Schwab JD, Schobel J, Werle SD, Furstberger A, Ikonomi N, Szekely R, et al. Perspective on mHealth concepts to ensure users' empowerment–from adverse event tracking for COVID-19 vaccinations to oncological treatment. IEEE Access 2021;9:83863-83875. [doi: 10.1109/access.2021.3087315]

70. Kestler AMR, Kühlwein SD, Kraus JM, Schwab JD, Szekely R, Thiam P, et al. Digitalization of adverse event management in oncology to improve treatment outcome-a prospective study protocol. PLoS One 2021;16(6):e0252493 [FREE Full text] [doi: 10.1371/journal.pone.0252493] [Medline: 34086740]
71. Kessels RPC. Patients' memory for medical information. J R Soc Med 2003 May;96(5):219-222 [FREE Full text] [doi: 10.1258/jrsm.96.5.219] [Medline: 12724430]
72. Schlee W, Pryss RC, Probst T, Schobel J, Bachmeier A, Reichert M, et al. Measuring the moment-to-moment variability of tinnitus: the TrackYourTinnitus smart phone app. Front Aging Neurosci 2016;8:294. [doi: 10.3389/fnagi.2016.00294] [Medline: 28018210]
73. Völkel G, Fürstberger A, Schwab JD, Werle SD, Ikonomi N, Gscheidmeier T, et al. Patient empowerment during the COVID-19 pandemic by ensuring safe and fast communication of test results: implementation and performance of a tracking system. J Med Internet Res 2021 Jun 07;23(6):e27348 [FREE Full text] [doi: 10.2196/27348] [Medline: 33999836]

Abbreviations

AI: artificial intelligence
CHW: community health worker
HCP: health care provider
mHealth: mobile health
WHO: World Health Organization

Edited by L Buis, A Mavragani; submitted 06.08.21; peer-reviewed by R Pryss, B Schreiweis; comments to author 24.11.21; revised version received 13.12.21; accepted 20.02.22; published 13.04.22.

Please cite as:
Hamberger M, Ikonomi N, Schwab JD, Werle SD, Fürstberger A, Kestler AMR, Holderried M, Kaisers UX, Steger F, Kestler HA
Interaction Empowerment in Mobile Health: Concepts, Challenges, and Perspectives
JMIR Mhealth Uhealth 2022;10(4):e32696
URL: https://mhealth.jmir.org/2022/4/e32696
doi:10.2196/32696
PMID:35416786

Viewpoint

Reimagining Connected Care in the Era of Digital Medicine

Devin M Mann[1,2], MD, MS; Katharine Lawrence[1,2], MD, MPH

[1]Department of Population Health, NYU Grossman School of Medicine, New York, NY, United States
[2]Medical Center Information Technology, NYU Langone Health, New York, NY, United States

Corresponding Author:
Devin M Mann, MD, MS
Department of Population Health
NYU Grossman School of Medicine
227 E 30th Street
New York, NY, 10016
United States
Phone: 1 2122639026
Email: devin.mann@nyulangone.org

Abstract

The COVID-19 pandemic accelerated the adoption of remote patient monitoring technology, which offers exciting opportunities for expanded connected care at a distance. However, while the mode of clinicians' interactions with patients and their health data has transformed, the larger framework of how we deliver care is still driven by a model of episodic care that does not facilitate this new frontier. Fully realizing a transformation to a system of continuous connected care augmented by remote monitoring technology will require a shift in clinicians' and health systems' approach to care delivery technology and its associated data volume and complexity. In this article, we present a solution that organizes and optimizes the interaction of automated technologies with human oversight, allowing for the maximal use of data-rich tools while preserving the pieces of medical care considered uniquely human. We review implications of this "augmented continuous connected care" model of remote patient monitoring for clinical practice and offer human-centered design-informed next steps to encourage innovation around these important issues.

(JMIR Mhealth Uhealth 2022;10(4):e34483) doi:10.2196/34483

KEYWORDS

health information technology; telehealth; remote patient monitoring; mobile health; mHealth; eHealth; digital health; innovation; process model; information technology; digital medicine; automation

The Growth of Remote Patient Monitoring: Accelerating the Transition From Episodic to Continuous Care

The COVID-19 pandemic accelerated the adoption of remote patient monitoring (RPM)—the use of ambulatory, noninvasive digital technology to capture and transmit patient data in real time for care delivery and disease management. The use of RPM technology during the pandemic has allowed for care at a distance in an age of unprecedented health uncertainty and disruption [1]. However, while the mode of clinicians' interactions with patients and their health data has transformed, the larger framework of *how we deliver care* has only incrementally shifted. Most health care is still delivered in episodes—synchronous moments of connection between clinicians and patients, mediated by discrete hospitalizations, office visits, or video and audio calls. But health is *not* episodic; it is a fundamental part of the human condition, experienced regularly and continuously, more akin to a utility (eg, energy, water, or even education) than a traditional professional service (eg, tax preparation). The friction in these competing visions of care has contributed to a fragmented, inconsistent health care delivery experience; it has also limited the health information technology resources, innovations, and capital needed to make the world of data-driven continuous care possible. RPM technologies have the potential to enable and accelerate a transition from episodic to continuous care. Here, we outline the current state of RPM, its challenges in clinical practice, and how a continuous connected care model can be organized based on technology-driven transitions that include not only RPM but also the larger world of digital health technologies (eg, telehealth, machine learning, and artificial intelligence [AI]).

The Tidal Wave of RPM Data

The rapid adoption of new digital health care technologies for remote health care provision has begun to dismantle the barrier between the clinic and the home. The influx of health data from

RPM devices has the potential to realize a new framework of medicine—not just of higher quality episodic medicine but of a continuous connected care model that reflects the true interaction of health care and the human experience. However, while more continuous and complete data from patients can facilitate this new world of care, its introduction into the health care landscape has also raised concerns [2]. RPM devices generate more data than a clinician is equipped to manage, and current health information technology systems are inadequate for the data curation and visualization required to effectively use these data to help patients [3]. At the same time, there are worries that simply increasing the *volume* of data about a patient does not translate to improved health outcomes, and that more understanding of the *quality* requirements of RPM is needed [4]. Finally, the increased burden on patients and clinicians to *always be connected*—whether to their health or their jobs—has raised concerns about medical overutilization, burnout, and excessive consumerism; even the term "remote patient monitoring" is problematic, conjuring images of invasive surveillance and control rather than supportive care [5,6].

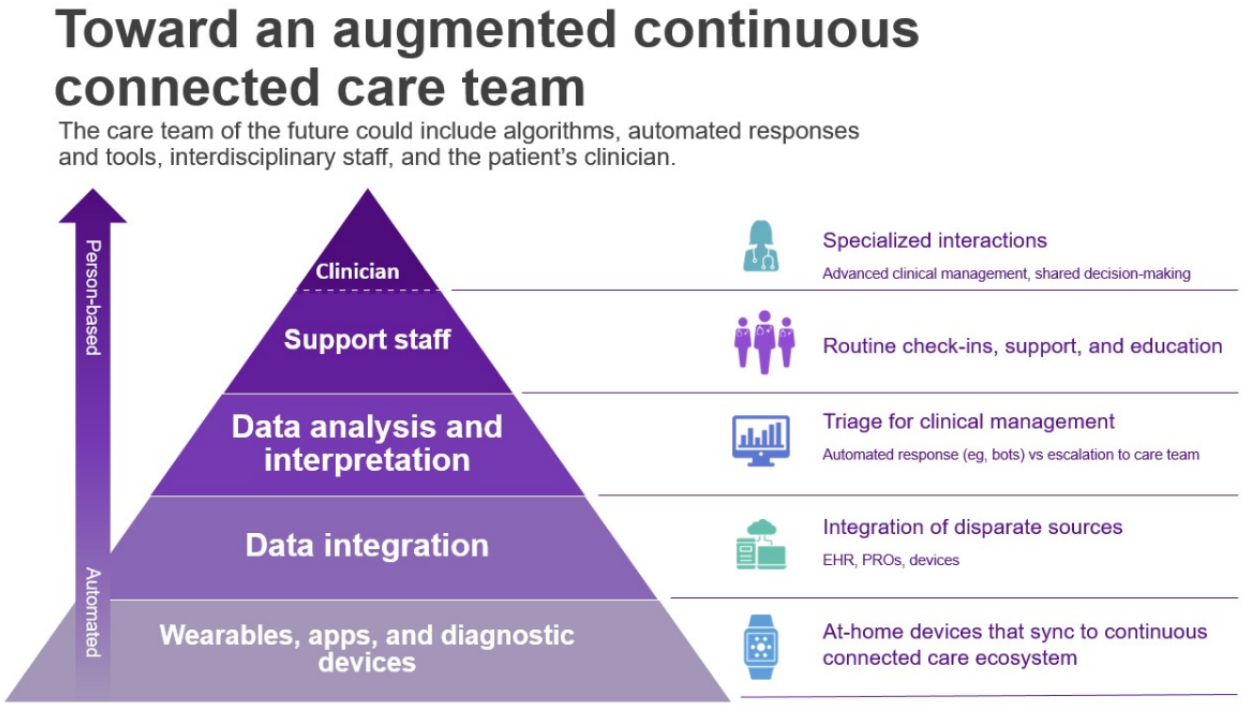
RPM and the Augmented Continuous Connected Care Pyramid

To address these challenges, health systems have taken several approaches to managing RPM in clinical practice; these include limiting digital access and services for patients, adding new staff, or implementing automated tools such as chatbots and AI models to manage the data and their clinical consequences [7,8]. Digital health vendors are increasingly offering (and payors are reimbursing) digital and virtual health services that include RPM devices with improving interoperability (leveraging standard interfaces) among vendors and electronic health records [9-11]. These new platforms enable more robust acute, chronic, and home hospital management [12-15]. Other digital health tools have been implemented that use AI to evaluate RPM data inputs and help boost the important signals buried within the noise. For example, DreaMed uses AI to sift through the mountains of RPM-generated continuous glucose monitoring and insulin pump data to make specific recommendations for insulin titrations [16]. Clinical staff are increasingly being tasked with managing the incoming RPM data streams and supporting providers in delivering a more continuous connected care experience. Ochsner's RPM driven hypertension program is a successful example of this new paradigm in practice [17,18]. Ultimately, a solution that optimizes the interaction of automated technologies with human oversight, both routine and specialized, will likely prevail, allowing for maximal use of data-rich technologies while preserving the pieces of medical care considered "uniquely" human.

Taken together, RPM tools and their implementation in health care systems can be viewed as an example of a novel pyramid of health care delivery, "augmented continuous connected care" (Figure 1). At the base, continuous connected care is built on an always "on," automated, and often passive, holistic health data capture layer. These diverse data points are integrated and standardized by an algorithmic (or a machine learning) layer that can "listen" to and interpret data and either respond autonomously or reduce noise and boost signals to generate more actionable insights for human interpretation. Decisions beyond the scope of this "digital clinician" are routed to the clinical team for management, enabling them to work to the top of their licenses and provide the parts of medical care that benefit most from a human touch—patient education, shared decision-making, and complex medical decision-making.

The benefits of RPM-enabled connected continuous care continue to emerge as the pandemic progresses and health care systems respond and evolve. Technologies that were once considered niche products for rural health or areas with resource or accessibility barriers have proven translatable to a variety of health care settings and contexts, enabling diverse populations to manage both acute and chronic conditions with improved information, safety, and convenience [19,20]. Shared management of RPM data, particularly after AI processing, is a more sustainable model for supporting continuous care and allows patients greater connectedness with their health care team.

Figure 1. Augmented continuous connected care pyramid. EHR: electronic health record; PRO: patient-reported outcomes.



Using Design Thinking to Identify and Overcome Challenges to RPM-Enabled Continuous Care

The impact of health care's shift to an augmented continuous connected care paradigm will be significant for patients, providers, and care delivery systems. Below, we frame key implications as a set of Design Thinking–informed "how might we" questions to encourage new thinking around these issues [21]. These questions are by no means exhaustive of the challenges facing the future of continuous connected care, but are designed to inspire health care leaders, designers, and clinicians to reimagine these and other key concerns related to RPM and the broader world of digital health technology.

1. How might we use RPM technologies to let patients easily share their "daily life data" without overwhelming their health care teams?
2. How might we change how patients and health care teams communicate to take advantage of continuous connected care?
3. How might we transform the health care workforce to manage this new type of high volume, daily life data?

These questions acknowledge both the potential and the challenges of using RPM and other digital health tools to evolve the world of health care. The answers to these questions require improvements in the design, usability, and interoperability of RPM tools to reduce the burden of patient work and empower equitable access to and use of this transformative technology. It will also require careful balancing of the benefits of continuous care with the risks of being "too connected." New approaches will be needed that help patients and clinicians stay updated on information and make actionable decisions on the key events in their health while also protecting themselves and their data.

Implications of Augmented Continuous Connected Care for Practitioners, Researchers, and Policy Makers

The rapid growth of RPM and the continuous connected care paradigm it enables requires new research and development to improve the devices, the platforms, integrated AI, interoperability, and the usability of RPM tools. New team structures, personnel, and workflows will need to be identified, tested, and disseminated to help health care systems and patients take full advantage of this enabling technology. Clinicians will need better EHR-integrated tools, training, and team-based support to manage patients using a mix of clinic and home-based data collection and leverage AI-based tools to highlight critical insights and actions. To support these new workflows, clinical delivery system leaders will need organizational and regulatory flexibility to modify health care workers' composition and scopes of work; such support may require investment in new roles such as RPM "navigators," community health workers, and others who manage data and devices in partnership with patients and clinicians. Policy makers will need to foster a regulatory environment and financial incentives that incentivizes RPM innovation while shepherding the industry to adhere to common standards and the free flow of data across systems.

RPM and other digital health technologies are ushering in a transition to a new and exciting model of care, from episodic to continuous and connected. However, there are many challenges facing this transition, and how we respond to them will shape the ultimate impact of not only these tools themselves but also our ability to use these technologies to improve health care experiences and outcomes.

Acknowledgments

The authors would like to thank their collaborators in the NYU Langone HiBIRD (Healthcare Innovation Bridging Research, Informatics, and Design Lab), the NYU Langone Health Medical Center Information Technology (MCIT) departments, and the MCIT FuturePractice. We would also like to thank Tiffany Martinez for her assistance in the manuscript preparation.

Conflicts of Interest

None declared.

References

1. Mann DM, Chen J, Chunara R, Testa PA, Nov O. COVID-19 transforms health care through telemedicine: Evidence from the field. J Am Med Inform Assoc 2020 Jul 01;27(7):1132-1135 [FREE Full text] [doi: 10.1093/jamia/ocaa072] [Medline: 32324855]
2. Houlding E, Mate KKV, Engler K, Ortiz-Paredes D, Pomey M, Cox J, et al. Barriers to Use of Remote Monitoring Technologies Used to Support Patients With COVID-19: Rapid Review. JMIR Mhealth Uhealth 2021 Apr 20;9(4):e24743 [FREE Full text] [doi: 10.2196/24743] [Medline: 33769943]
3. Kruse CS, Goswamy R, Raval Y, Marawi S. Challenges and Opportunities of Big Data in Health Care: A Systematic Review. JMIR Med Inform 2016 Nov 21;4(4):e38 [FREE Full text] [doi: 10.2196/medinform.5359] [Medline: 27872036]
4. Adibuzzaman M, DeLaurentis P, Hill J, Benneyworth BD. Big data in healthcare - the promises, challenges and opportunities from a research perspective: A case study with a model database. AMIA Annu Symp Proc 2017;2017:384-392 [FREE Full text] [Medline: 29854102]
5. Latimer T, Roscamp J, Papanikitas A. Patient-centredness and consumerism in healthcare: an ideological mess. J R Soc Med 2017 Nov;110(11):425-427 [FREE Full text] [doi: 10.1177/0141076817731905] [Medline: 28949269]
6. Gerke S, Shachar C, Chai PR, Cohen IG. Regulatory, safety, and privacy concerns of home monitoring technologies during COVID-19. Nat Med 2020 Aug;26(8):1176-1182 [FREE Full text] [doi: 10.1038/s41591-020-0994-1] [Medline: 32770164]
7. Jadczyk T, Wojakowski W, Tendera M, Henry TD, Egnaczyk G, Shreenivas S. Artificial Intelligence Can Improve Patient Management at the Time of a Pandemic: The Role of Voice Technology. J Med Internet Res 2021 May 25;23(5):e22959 [FREE Full text] [doi: 10.2196/22959] [Medline: 33999834]
8. Coffey JD, Christopherson LA, Glasgow AE, Pearson KK, Brown JK, Gathje SR, et al. Implementation of a multisite, interdisciplinary remote patient monitoring program for ambulatory management of patients with COVID-19. NPJ Digit Med 2021 Aug 13;4(1):123 [FREE Full text] [doi: 10.1038/s41746-021-00490-9] [Medline: 34389787]
9. Omron Remote Patient Monitoring Products. Omron Healthcare. URL: https://omronhealthcare.com/ [accessed 2022-01-12]
10. TytoHome Telehealth Equipment. Tytocare. URL: https://www.tytocare.com/professionals/products/tytohome/ [accessed 2022-01-12]
11. Contessa Health. URL: https://contessahealth.com/solutions/ [accessed 2022-01-12]
12. Brody J. A Pandemic Benfit: The Expansion of Telemedicine. The New York Times. 2020 May 11. URL: https://www.nytimes.com/2020/05/11/well/live/coronavirus-telemedicine-telehealth.html?smid=url-share [accessed 2022-01-12]
13. Henderson K. Accelerating a future where health care is closer to home. Optum. 2021. URL: https://www.optum.com/business/resources/health-care-tech-blog/home-health-care.html [accessed 2022-03-30]
14. Lishman LJ. Remote patient monitoring: the future of healthcare? PharmaTimes. 2021 Mar 22. URL: https://www.pharmatimes.com/web_exclusives/Remote_patient_monitoring_the_future_of_healthcare_1365643 [accessed 2022-01-12]
15. Crossen S, Romero C, Reggiardo A, Michel J, Glaser N. Feasibility and Impact of Remote Glucose Monitoring Among Patients With Newly Diagnosed Type 1 Diabetes: Single-Center Pilot Study. JMIR Diabetes 2022 Jan 17;7(1):e33639 [FREE Full text] [doi: 10.2196/33639] [Medline: 35037887]
16. Nimri R, Battelino T, Laffel LM, Slover RH, Schatz D, Weinzimer SA, NextDREAM Consortium. Insulin dose optimization using an automated artificial intelligence-based decision support system in youths with type 1 diabetes. Nat Med 2020 Sep;26(9):1380-1384. [doi: 10.1038/s41591-020-1045-7] [Medline: 32908282]
17. Milani RV, Lavie CJ, Wilt JK, Bober RM, Ventura HO. New Concepts in Hypertension Management: A Population-Based Perspective. Prog Cardiovasc Dis 2016;59(3):289-294. [doi: 10.1016/j.pcad.2016.09.005] [Medline: 27693860]
18. Milani RV, Lavie CJ, Bober RM, Milani AR, Ventura HO. Improving Hypertension Control and Patient Engagement Using Digital Tools. Am J Med 2017 Jan;130(1):14-20. [doi: 10.1016/j.amjmed.2016.07.029] [Medline: 27591179]
19. Hutchings OR, Dearing C, Jagers D, Shaw MJ, Raffan F, Jones A, et al. Virtual Health Care for Community Management of Patients With COVID-19 in Australia: Observational Cohort Study. J Med Internet Res 2021 Mar 09;23(3):e21064 [FREE Full text] [doi: 10.2196/21064] [Medline: 33687341]
20. Annis T, Pleasants S, Hultman G, Lindemann E, Thompson JA, Billecke S, et al. Rapid implementation of a COVID-19 remote patient monitoring program. J Am Med Inform Assoc 2020 Aug 01;27(8):1326-1330 [FREE Full text] [doi: 10.1093/jamia/ocaa097] [Medline: 32392280]

21. Melles M, Albayrak A, Goossens R. Innovating health care: key characteristics of human-centered design. Int J Qual Health Care 2021 Jan 12;33(Supplement_1):37-44 [FREE Full text] [doi: 10.1093/intqhc/mzaa127] [Medline: 33068104]

Abbreviations

AI: artificial intelligence
RPM: remote patient monitoring

Edited by L Buis; submitted 25.10.21; peer-reviewed by J Marquard, G Davlyatov, B Chaudhry, L Chirchir; comments to author 21.12.21; revised version received 18.01.22; accepted 11.03.22; published 15.04.22.

Please cite as:
Mann DM, Lawrence K
Reimagining Connected Care in the Era of Digital Medicine
JMIR Mhealth Uhealth 2022;10(4):e34483
URL: https://mhealth.jmir.org/2022/4/e34483
doi:10.2196/34483
PMID:35436238

Viewpoint

SciKit Digital Health: Python Package for Streamlined Wearable Inertial Sensor Data Processing

Lukas Adamowicz[1], MSc; Yiorgos Christakis[1], MSc; Matthew D Czech[1], PhD; Tomasz Adamusiak[1], MD, PhD

Digital Medicine and Translational Imaging, Pfizer Inc, Cambridge, MA, United States

Corresponding Author:
Lukas Adamowicz, MSc
Digital Medicine and Translational Imaging
Pfizer Inc
610 Main Street
Cambridge, MA, 02139
United States
Phone: 1 802 324 1829
Email: lukas.adamowicz@pfizer.com

Abstract

Wearable inertial sensors are providing enhanced insight into patient mobility and health. Significant research efforts have focused on wearable algorithm design and deployment in both research and clinical settings; however, open-source, general-purpose software tools for processing various activities of daily living are relatively scarce. Furthermore, few studies include code for replication or off-the-shelf software packages. In this work, we introduce SciKit Digital Health (SKDH), a Python software package (Python Software Foundation) containing various algorithms for deriving clinical features of gait, sit to stand, physical activity, and sleep, wrapped in an easily extensible framework. SKDH combines data ingestion, preprocessing, and data analysis methods geared toward modern data science workflows and streamlines the generation of digital endpoints in "good practice" environments by combining all the necessary data processing steps in a single pipeline. Our package simplifies the construction of new data processing pipelines and promotes reproducibility by following a convention over configuration approach, standardizing most settings on physiologically reasonable defaults in healthy adult populations or those with mild impairment. SKDH is open source, as well as free to use and extend under a permissive Massachusetts Institute of Technology license, and is available from GitHub (PfizerRD/scikit-digital-health), the Python Package Index, and the conda-forge channel of Anaconda.

(JMIR Mhealth Uhealth 2022;10(4):e36762) doi:10.2196/36762

KEYWORDS

wearable sensors; digital medicine; gait analysis; human movement analysis; digital biomarkers; uHealth; wearable; sensor; gait; movement; mobility; physical activity; sleep; Python; coding; open source; software package; algorithm; machine learning; data science; computer programming

Introduction

Wearable inertial sensors have enabled huge leaps forward in the ability to quantify and derive actionable insights from patient mobility and at-home health. Algorithm development and deployment in both research and clinical studies have been a focus of many research efforts. For example, gait monitoring using wearables has evolved from algorithm design using minimal sensors for the purpose of minimizing patient burden [1-5] to at-home deployment and remote monitoring of free-living activity [6-8]. Remote patient monitoring has a high intrinsic value, as previous work has suggested. At-home values may be less influenced by observer effects [8,9] and may facilitate enhanced group differentiation [8-10].

While lumbar-mounted sensors are appealing for capturing bilateral gait and other lower body activities such as sit-to-stand transfers, wrist sensors are also desirable as they can be integrated into watches or watch-like packages and offer lower subject burden. Sleep and physical activity monitoring, which typically use a wrist-based sensor, are also among extensively researched areas [11-18]. Sleep and physical activity research have been aided by the availability of an open-source, freely available code package, GGIR [19]. GGIR is a collection of algorithms for activity and sleep research, written in R, and includes code to ingest, calibrate, and detect sleep and activity level from raw acceleration data. GGIR allows researchers to study patient symptoms with limited programming expertise and has been evaluated in over 90 peer-reviewed journal publications [19].

The availability of GGIR is in stark contrast with many other published works in this area. Relatively few works include any code for experiment replication, and even fewer include easy-to-use or "off-the-shelf" code packages, despite the ease of sharing through public code repositories such as GitHub. Our group has made an effort to release several implementations from existing research or new algorithms for gait [8], sleep [20], and sit to stand [10]. However, open-source packages to date are fairly disparate and require additional steps for data ingestion and preprocessing. Other options include the Digital Biomarker Discovery Pipeline [21], a partial set of tools with the goal of enhancing data inspection, cleaning, and processing to enable digital biomarker discovery. However, it is composed of separate modules with iPython notebooks instead of Python libraries, and currently the project seems dormant (the last update was on November 3, 2020). Open-source GENEActiv R macros also exist, even though they are specific to GENEActiv files and would require custom modification to ingest data from other devices.

The lack of open-source, general-purpose algorithms for the processing of the various base activities of daily living is a significant gap in the field. By addressing this limitation, we hope to advance human activity recognition research in two important ways: (1) lowering the requirements for analyzing longitudinal data and (2) providing a baseline set of algorithm implementations for the community. Additionally, given the ease of sharing code, we hope to encourage the practice of sharing code with publications—an approach that should be adopted from other areas such as machine or deep learning research and encouraged by the National Institutes of Health.

In this paper, we present a new Python package, SciKit Digital Health (SKDH), to address the lack of open-source, general-purpose algorithms for monitoring digital health. SKDH contains algorithms for various measures of human activity recognition and streamlines the data ingestion, preprocessing, and data analysis steps. While the underlying algorithms themselves are not necessarily novel work, the novelty and utility of this work is the collection of common mobility and activity algorithms under a common framework that is being released open source. SKDH aims to address the shortcomings in available, existing codebases by (1) being easily usable with minimal interaction required from end users; (2) being tightly integrated so that different processing modules can be easily chained together, allowing multiple preprocessing and analysis steps in the same pipeline; and (3) being free and open source.

Methods

SciKit Digital Health

SKDH is a Python 3 package that contains algorithms for gait, sit to stand, activity level, and sleep. Additionally, it contains various preprocessing methods such as accelerometer calibration; wear detection; and binary file data readers for the GENEActiv, Axivity, and ActiGraph sensors. Individual algorithms or steps are built around an extensible process class ("BaseProcess"), which are chained together as needed in a pipeline structure. The BaseProcess class abstracts various setup tasks and standardized functions that allow for subclasses to function properly and in sequence in the SKDH framework. This allows the end user to easily link steps together, as shown in Textbox 1.

SKDH also contains various common utility functions (eg, moving mean, standard deviation with arbitrary window length, and skip values) and a suite of features for signal processing and feature generation for machine learning, written in C or Fortran, to reduce computation time (Table 1).

A more comprehensive example that shows how SKDH base classes can be extended and easily integrated into an SKDH pipeline is shown in Figure 1.

Additionally, to simulate a realistic processing scenario, the data was windowed over 3-second windows (150 samples) with 50% overlap, and the computation was run again.

Textbox 1. Example script that will (1) import data from a GENEActiv bin file, (2) calibrate the accelerometer so that still periods measure 1 g, and (3) run gait processing to generate gait endpoints.

```
import skdh

pipeline=skdh.Pipeline()

pipeline.add(skdh.io.ReadBin())

pipeline.add(skdh.preprocessing.CalibrateAccelerometer())

pipeline.add(skdh.gait.Gait())

pipeline.run(file="example_geneactiv_file.bin")
```

Table 1. Mean (SD) processing times in milliseconds on a representative array of random[a] data.

Feature	100,000 3 array[b]			Windowed: 3s, 50% overlap[c]		
	Original[d] (ms), mean (SD)	SciKit Digital Health (ms), mean (SD)	Factor	Original (ms), mean (SD)	SKDH (ms), mean (SD)	Factor
Signal entropy	12.7 (0.31)	1.53 (0.03)	8.3	3008 (88.7)	3.89 (0.21)	792
Jerk metric	22.6 (1.86)	0.05 (0.02)	45	2720 (80.8)	0.97 (0.07)	2810
Spectral arc length	1005 (24.7)	197 (3.70)	5.3	3340 (102)	115 (3.74)	29

[a]NumPy.random.default_rng().standard normal.

[b]Produces 3 values for the feature.

[c]1332 resulting windows. Original runs 3 separate data frames (shape (150, 1332)), one for each XYZ axis. SKDH features run on, full shape (1332, 150, 3) array.

[d]Implemented with NumPy for Pandas input.

Figure 1. A custom class for reading a file from a new device is first created as a subclass of SciKit Digital Health's (SKDH) "BaseProcess" that allows it to be easily inserted into a SKDH pipeline. Note that SKDH will automatically save results from the default sleep and activity analyses to the specified files.

```
# skdh_extension.py
from such import BaseProcess

class ReadNewDeviceFile(BaseProcess):
    def __init__(self, bases=None, periods=None):
        super().__init__(bases=bases, periods=periods)
        # custom __init__

    def predict(self, file=None, **kwargs):
        super().predict(
            expect_days=False, expect_wear=False, file=file, **kwargs)
        # New device file reader contents here
# process_new_device_file.py
import skdh
from skdh_extension import ReadNewDeviceFile

pipeline = skdh.Pipeline()
pipeline.add(ReadNewDeviceFile(bases=[3], periods=[18]))
pipeline.add(skdh.preprocessing.CalibrateAccelerometer())
pipeline.add(skdh.preprocessing.DetectWear())
pipeline.add(
    skdh.sleep.Sleep(),
    save_file="sleep_results.csv",
    plot_file="sleep_visualization.pdf",
)
pipeline.add(
    skdh.activity.ActivityLevelClassification(),
    save_file="activity_results.csv",
    plot_file="activity_visualization.pdf",
)
pipeline.run(file="new_device_file.abc")
```

Gait

The gait algorithm uses the inverted pendulum model of gait to extract bilateral gait endpoints from acceleration data collected from a lumbar-mounted wearable inertial sensor [1-4]. In general, gait bouts during free-living data are first detected using a gradient boosted tree classifier [22]. For in-lab data in which the time periods of gait are known, the gait classification step can be skipped. Wavelet transforms are then used to detect initial and final contact events for each foot from the vertical acceleration signal [4]. With these contact events, all temporal endpoints (eg, stride time, double support, etc) are computed.

In order to obtain spatial metrics (eg, stride length, gait speed, etc), an inverted pendulum model [1] is used, requiring only the participant's height in addition to the vertical acceleration signal.

The implementation is very similar to that of our previously released GaitPy package [8], updated to fit into the SKDH architecture, with a few key algorithm additions and updates. Notably, the classifier for gait bouts during at-home periods has been updated, using the training data from 4 additional studies to gain a better breadth of nongait activities. These studies are "the daily life activities" [23], "the long term movement monitoring database" [24], "the University of Southern California human activity dataset" [25], and "a Parkinson's disease study" [26].

In the original GaitPy wavelet transform implementation, a fixed scale was used. However, recent research shows that the scale can be better optimized by matching it to the step frequency [5], and this relationship was added as an optional toggle. Finally, additional asymmetry endpoints were added, including but not limited to the gait symmetry index [27-29], step and stride regularity [3,28], and intrastep and intrastride covariance [28].

Sit to Stand

The sit-to-stand algorithm is identical to what was released in Sit2StandPy [10], though integrated into the SKDH framework. It uses acceleration data from a lumbar-mounted device to identify sit-to-stand transfers in both in-lab and free-living environments. The sit-to-stand algorithm is a heuristic algorithm, which functions by identifying possible sit-to-stand locations using a wavelet transform and acceleration filtering. With possible locations identified, a series of quality checks and rules are imposed to determine whether the transfer is valid or not. Validation for the algorithm was previously presented using data from patients with Parkinson's disease and healthy adults [10].

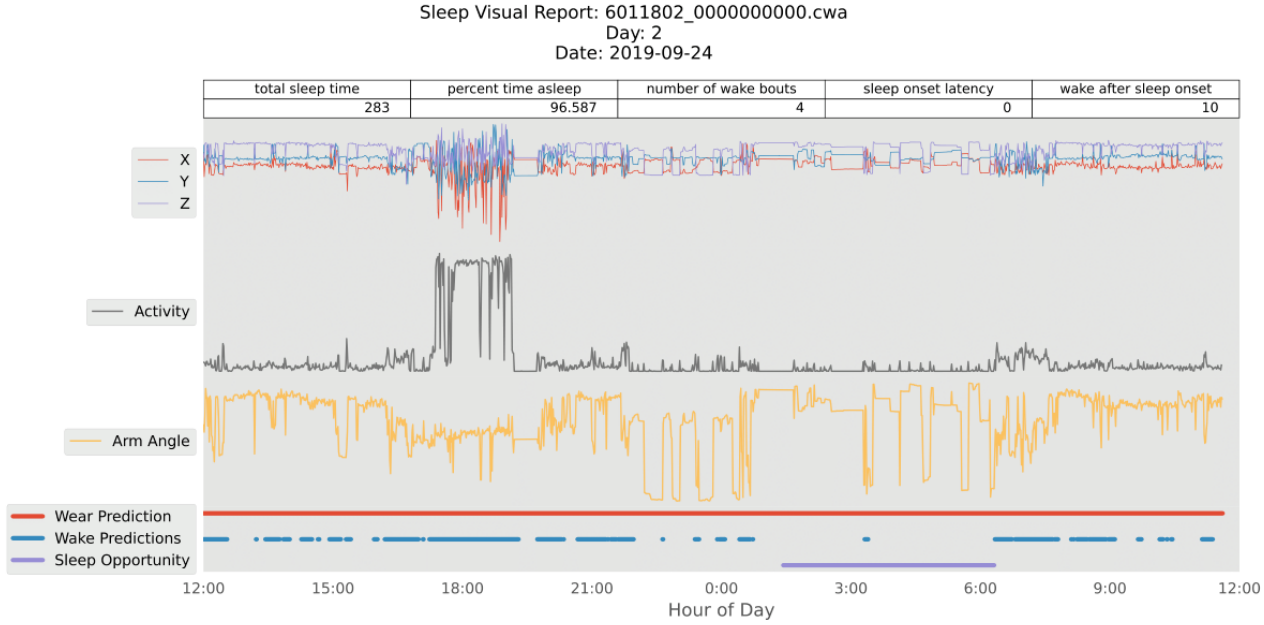
Sleep

The sleep algorithm in SKDH was originally presented in the Python package, SleepPy [20-30], and here, it was adapted into the SKDH framework. This algorithm was originally based on the one implemented in GGIR [31]. It is intended for use on the acceleration data from the wrist, even though it will also take advantage of near-body temperature data, if available, to significantly improve sleep-specific, on-body detection. The algorithm first determines 1 sleep opportunity window per day (noon to noon) using a series of moving mean and median filters. During this period, bouts of sleep and wake are determined by computing the activity index [32] of the acceleration data and then applying a heuristic scoring algorithm [33]. Sleep endpoints are then calculated, including but not limited to wake after sleep onset, total sleep time, as well as sleep and wake transition probabilities [34,35]. If desired, a per-day sleep plot can be produced as well for visual inspection.

Activity

The activity algorithm seeks to provide similar outputs to previously published research [15,16,18] such as time spent in sedentary, light, moderate, or vigorous activity levels. Wrist-based triaxial acceleration is windowed into 5-second blocks, and the mean is taken. By default, the value of gravity is subtracted to obtain the Euclidean norm minus one (ENMO). These ENMO values are then used to threshold into different activity levels with different provided base options derived from the literature [13,15,16,18]. These periods of time in different activity levels can also be rescored to obtain bouts of consistent activity level [14,18,36]. Finally, recent work has also proposed alternative methods of accessing activity level by quantifying the decline in the time spent in increasing activity magnitude [17]; this analysis is also included in the activity endpoints. Similar to the sleep plot, a per-day activity plot can be saved if desired, showing the acceleration, activity, activity level, and wear traces, as seen in Figure 2.

Figure 2. A sample sleep plot as produced by SciKit Digital Health, showing a single night from test data.



Research Applications

Use cases for research applications are widely varied and cover a broad spectrum of research topics in the relevant fields. First and foremost, SKDH provides a quick and easy-to-use tool to generate activity and mobility endpoints with limited adjustment and setup required on the part of researchers or clinicians. Since default parameter values for algorithms are set to physiological defaults, SKDH would provide an "off-the-shelf" experience when the research goal is endpoint assessment or comparison.

However, the adjustable algorithm parameters also allow for a more nuanced approach if the research goal is instead exploring the algorithms themselves. Along with this, as the code is open source, researchers are also able to use SKDH as a starting point and add functionality or improvements as they need for their work.

These utilization strategies for SKDH in research on gait, sit to stand, activity, and sleep lead to a broad range of applications for SKDH in research. On top of this, many of the additional utility or feature generation capabilities present in SKDH are useful outside the context of these activities as well, for initial data exploration or even just for ingestion of data from sensor binary file formats.

Validation

Validation of algorithm implementation is critical to ensure that the generated results match the expected values and provide actionable insight. For SKDH, validation is an ongoing effort with the different modules having different levels of validation, even though all the individual algorithms were validated in their original publications. Validation for the sit-to-stand module included in SKDH was presented previously [10], and the algorithm implementation remained exactly the same. The gait and the sleep modules had previous implementations validated in previous publications (gait in a study by Czech et al [8], and sleep in a study by Mahadevan et al [20]), even though there are implementation differences and algorithm additions in SKDH. Internal validation of the gait module showed a higher agreement and tighter ranges of intraclass correlation coefficients compared with the previous versions of the gait implementation (results not shown). The activity module has also shown excellent agreement in internal comparisons to GENEActiv macros and GGIR (results not shown).

Acknowledgments

This work was sponsored by Pfizer Inc.

Data Availability

The source code is available under the Massachusetts Institute of Technology license from GitHub (PfizerRD/scikit-digital-health). Installable packages will be available on the Python Package Index and the conda-forge channel of Anaconda.

Authors' Contributions

LA, MDC, and YC contributed to SKDH. LA and TA contributed to writing the paper and project supervision. All authors reviewed the paper and provided approval.

Conflicts of Interest

None declared.

References

1. Zijlstra W, Hof A. Displacement of the pelvis during human walking: experimental data and model predictions. Gait & Posture 1997 Dec;6(3):249-262. [doi: 10.1016/s0966-6362(97)00021-0]
2. Zijlstra W. Assessment of spatio-temporal parameters during unconstrained walking. Eur J Appl Physiol 2004 Jun;92(1-2):39-44. [doi: 10.1007/s00421-004-1041-5] [Medline: 14985994]
3. Moe-Nilssen R, Helbostad JL. Estimation of gait cycle characteristics by trunk accelerometry. J Biomech 2004 Jan;37(1):121-126. [doi: 10.1016/s0021-9290(03)00233-1]
4. McCamley J, Donati M, Grimpampi E, Mazzà C. An enhanced estimate of initial contact and final contact instants of time using lower trunk inertial sensor data. Gait & Posture 2012 Jun;36(2):316-318. [doi: 10.1016/j.gaitpost.2012.02.019] [Medline: 22465705]
5. Caramia C, De Marchis C, Schmid M. Optimizing the Scale of a Wavelet-Based Method for the Detection of Gait Events from a Waist-Mounted Accelerometer under Different Walking Speeds. Sensors (Basel) 2019 Apr 19;19(8):1869 [FREE Full text] [doi: 10.3390/s19081869] [Medline: 31010114]
6. Del Din S, Godfrey A, Galna B, Lord S, Rochester L. Free-living gait characteristics in ageing and Parkinson's disease: impact of environment and ambulatory bout length. J Neuroeng Rehabil 2016 May 12;13(1):46 [FREE Full text] [doi: 10.1186/s12984-016-0154-5] [Medline: 27175731]
7. Del Din S, Galna B, Godfrey A, Bekkers E, Pelosin E, Nieuwhof F, et al. Analysis of Free-Living Gait in Older Adults With and Without Parkinson's Disease and With and Without a History of Falls: Identifying Generic and Disease-Specific

Characteristics. J Gerontol A Biol Sci Med Sci 2019 Mar 14;74(4):500-506 [FREE Full text] [doi: 10.1093/gerona/glx254] [Medline: 29300849]
8. Czech MD, Psaltos D, Zhang H, Adamusiak T, Calicchio M, Kelekar A, et al. Age and environment-related differences in gait in healthy adults using wearables. NPJ Digit Med 2020 Sep 30;3:127 [FREE Full text] [doi: 10.1038/s41746-020-00334-y] [Medline: 33083562]
9. Takayanagi N, Sudo M, Yamashiro Y, Lee S, Kobayashi Y, Niki Y, et al. Relationship between Daily and In-laboratory Gait Speed among Healthy Community-dwelling Older Adults. Sci Rep 2019 Mar 05;9(1):3496 [FREE Full text] [doi: 10.1038/s41598-019-39695-0] [Medline: 30837520]
10. Adamowicz L, Karahanoglu FI, Cicalo C, Zhang H, Demanuele C, Santamaria M, et al. Assessment of Sit-to-Stand Transfers during Daily Life Using an Accelerometer on the Lower Back. Sensors (Basel) 2020 Nov 19;20(22):6618 [FREE Full text] [doi: 10.3390/s20226618] [Medline: 33228035]
11. van Hees VT, Sabia S, Anderson KN, Denton SJ, Oliver J, Catt M, et al. A Novel, Open Access Method to Assess Sleep Duration Using a Wrist-Worn Accelerometer. PLoS One 2015;10(11):e0142533 [FREE Full text] [doi: 10.1371/journal.pone.0142533] [Medline: 26569414]
12. van Hees VT, Gorzelniak L, Dean León EC, Eder M, Pias M, Taherian S, et al. Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. PLoS One 2013 Apr 23;8(4):e61691 [FREE Full text] [doi: 10.1371/journal.pone.0061691] [Medline: 23626718]
13. Phillips LRS, Parfitt G, Rowlands AV. Calibration of the GENEA accelerometer for assessment of physical activity intensity in children. J Sci Med Sport 2013 Mar;16(2):124-128. [doi: 10.1016/j.jsams.2012.05.013] [Medline: 22770768]
14. da Silva IC, van Hees VT, Ramires V, Knuth A, Bielemann R, Ekelund U, et al. Physical activity levels in three Brazilian birth cohorts as assessed with raw triaxial wrist accelerometry. Int J Epidemiol 2014 Dec;43(6):1959-1968 [FREE Full text] [doi: 10.1093/ije/dyu203] [Medline: 25361583]
15. Hildebrand M, van Hees VT, Hansen BH, Ekelund U. Age group comparability of raw accelerometer output from wrist- and hip-worn monitors. Med Sci Sports Exerc 2014 Sep;46(9):1816-1824. [doi: 10.1249/MSS.0000000000000289] [Medline: 24887173]
16. Hildebrand M, Hansen BH, van Hees VT, Ekelund U. Evaluation of raw acceleration sedentary thresholds in children and adults. Scand J Med Sci Sports 2017 Dec;27(12):1814-1823. [doi: 10.1111/sms.12795] [Medline: 27878845]
17. Rowlands A, Edwardson C, Davies M, Khunti K, Harrington D, Yates T. Beyond Cut Points: Accelerometer Metrics that Capture the Physical Activity Profile. Med Sci Sports Exerc 2018 Jun;50(6):1323-1332 [FREE Full text] [doi: 10.1249/MSS.0000000000001561] [Medline: 29360664]
18. Migueles JH, Cadenas-Sanchez C, Rowlands AV, Henriksson P, Shiroma EJ, Acosta FM, et al. Comparability of accelerometer signal aggregation metrics across placements and dominant wrist cut points for the assessment of physical activity in adults. Sci Rep 2019 Dec 03;9(1):18235 [FREE Full text] [doi: 10.1038/s41598-019-54267-y] [Medline: 31796778]
19. Migueles JH, Rowlands AV, Huber F, Sabia S. GGIR: a research community–driven open source r package for generating physical activity and sleep outcomes from multi-day raw accelerometer data. JMPB 2019 Sep 23:188-196 [FREE Full text] [doi: 10.1123/jmpb.2018-0063]
20. Mahadevan N, Christakis Y, Di J, Bruno J, Zhang Y, Dorsey ER, et al. Development of digital measures for nighttime scratch and sleep using wrist-worn wearable devices. NPJ Digit Med 2021 Mar 03;4(1):42 [FREE Full text] [doi: 10.1038/s41746-021-00402-x] [Medline: 33658610]
21. Bent B, Wang K, Grzesiak E, Jiang C, Qi Y, Jiang Y, et al. The digital biomarker discovery pipeline: An open-source software platform for the development of digital biomarkers using mHealth and wearables data. J Clin Transl Sci 2020 Jul 14;5(1):e19 [FREE Full text] [doi: 10.1017/cts.2020.511] [Medline: 33948242]
22. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: a highly efficient gradient boosting decision tree. 2017 Presented at: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems; Dec 4-9; Long Beach, CA p. 3149-3157 URL: https://papers.nips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html
23. Leutheuser H, Schuldhaus D, Eskofier B. Hierarchical, multi-sensor based classification of daily life activities: comparison with state-of-the-art algorithms using a benchmark dataset. PLoS One 2013;8(10):e75196 [FREE Full text] [doi: 10.1371/journal.pone.0075196] [Medline: 24130686]
24. Ihlen EAF, Weiss A, Helbostad JL, Hausdorff JM. The Discriminant Value of Phase-Dependent Local Dynamic Stability of Daily Life Walking in Older Adult Community-Dwelling Fallers and Nonfallers. Biomed Res Int 2015;2015:402596 [FREE Full text] [doi: 10.1155/2015/402596] [Medline: 26491669]
25. Zhang M, Sawchuk A. USC-HAD: a daily activity dataset for ubiquitous activity recognition using wearable sensors. 2012 Presented at: Ubicomp '12: The 2012 ACM Conference on Ubiquitous Computing; Sept 5-8; Pittsburgh, PA p. 1036-1043. [doi: 10.1145/2370216.2370438]
26. Erb MK, Karlin DR, Ho BK, Thomas KC, Parisi F, Vergara-Diaz GP, et al. mHealth and wearable technology should replace motor diaries to track motor fluctuations in Parkinson's disease. NPJ Digit Med 2020;3:6 [FREE Full text] [doi: 10.1038/s41746-019-0214-x] [Medline: 31970291]

27. Zhang W, Smuck M, Legault C, Ith MA, Muaremi A, Aminian K. Gait Symmetry Assessment with a Low Back 3D Accelerometer in Post-Stroke Patients. Sensors 2018 Oct 03;18(10):3322 [FREE Full text] [doi: 10.3390/s18103322] [Medline: 30282947]
28. Buckley C, Micó-Amigo ME, Dunne-Willows M, Godfrey A, Hickey A, Lord S, et al. Gait Asymmetry Post-Stroke: Determining Valid and Reliable Methods Using a Single Accelerometer Located on the Trunk. Sensors (Basel) 2019 Dec 19;20(1):37 [FREE Full text] [doi: 10.3390/s20010037] [Medline: 31861630]
29. von Schroeder HP, Coutts RD, Lyden PD, Billings E, Nickel VL. Gait parameters following stroke: a practical assessment. J Rehabil Res Dev 1995 Feb;32(1):25-31. [Medline: 7760264]
30. Christakis Y, Mahadevan N, Patel S. SleepPy: A python package for sleep analysis from accelerometer data. JOSS 2019 Dec;4(44):1663. [doi: 10.21105/joss.01663]
31. van Hees VT, Sabia S, Jones SE, Wood AR, Anderson KN, Kivimäki M, et al. Estimating sleep parameters using an accelerometer without sleep diary. Sci Rep 2018 Aug 28;8(1):12975 [FREE Full text] [doi: 10.1038/s41598-018-31266-z] [Medline: 30154500]
32. Bai J, Di C, Xiao L, Evenson KR, LaCroix AZ, Crainiceanu CM, et al. An Activity Index for Raw Accelerometry Data and Its Comparison with Other Activity Metrics. PLoS One 2016;11(8):e0160644 [FREE Full text] [doi: 10.1371/journal.pone.0160644] [Medline: 27513333]
33. Cole R, Kripke D, Gruen W, Mullaney D, Gillin J. Automatic sleep/wake identification from wrist activity. Sleep 1992 Oct;15(5):461-469. [doi: 10.1093/sleep/15.5.461] [Medline: 1455130]
34. Di J, Leroux A, Urbanek J, Varadhan R, Spira A, Schrack J, et al. Patterns of sedentary and active time accumulation are associated with mortality in US adults: The NHANES study. BioRxiv Preprint posted online Aug 31, 2017. [doi: 10.1101/182337]
35. Wanigatunga AA, Di J, Zipunnikov V, Urbanek JK, Kuo P, Simonsick EM, et al. Association of Total Daily Physical Activity and Fragmented Physical Activity With Mortality in Older Adults. JAMA Netw Open 2019 Oct 02;2(10):e1912352 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.12352] [Medline: 31577355]
36. Sabia S, van Hees VT, Shipley MJ, Trenell MI, Hagger-Johnson G, Elbaz A, et al. Association between questionnaire- and accelerometer-assessed physical activity: the role of sociodemographic factors. Am J Epidemiol 2014 Mar 15;179(6):781-790 [FREE Full text] [doi: 10.1093/aje/kwt330] [Medline: 24500862]

Abbreviations

ENMO: Euclidean norm minus one
SKDH: Scikit Digital Health

Edited by L Buis; submitted 24.01.22; peer-reviewed by C Baxter, R Patel, V Gupta; comments to author 22.03.22; revised version received 28.03.22; accepted 29.03.22; published 21.04.22.

Please cite as:
Adamowicz L, Christakis Y, Czech MD, Adamusiak T
SciKit Digital Health: Python Package for Streamlined Wearable Inertial Sensor Data Processing
JMIR Mhealth Uhealth 2022;10(4):e36762
URL: https://mhealth.jmir.org/2022/4/e36762
doi:10.2196/36762
PMID:35353039

Viewpoint

Developing a Smart Home Technology Innovation for People With Physical and Mental Health Problems: Considerations and Recommendations

Cheryl Forchuk[1,2*], RN, PhD; Jonathan Serrato[1*], BSc, MSc; Daniel Lizotte[3*], PhD; Rupinder Mann[4*], BSc, BESc, PEng, PMP; Gavin Taylor[5*]; Sara Husni[1*], BHSc

[1]Mental Health Nursing Research Alliance, Parkwood Institute, Lawson Health Research Institute, London, ON, Canada
[2]Arthur Labatt Family School of Nursing, Western University, London, ON, Canada
[3]Department of Computer Science, Western University, London, ON, Canada
[4]Research Informatics Department, Lawson Health Research Institute, London, ON, Canada
[5]Information Technology Services, London Health Sciences Centre, London, ON, Canada
[*]all authors contributed equally

Corresponding Author:
Jonathan Serrato, BSc, MSc
Mental Health Nursing Research Alliance
Parkwood Institute
Lawson Health Research Institute
550 Wellington Road S
London, ON, N6C 0A7
Canada
Phone: 1 519 685 8500 ext 75802
Email: jonathan.serrato@lhsc.on.ca

Abstract

Smart home technologies present an unprecedented opportunity to improve health and health care by providing greater communication and connectivity with services and care providers and by supporting the daily activities of people managing both mental and physical health problems. Based on our experience from conducting smart technology health studies, including a smart home intervention, we provide guidance on developing and implementing such interventions. First, we describe the need for an overarching principle of security and privacy that must be attended to in all aspects of such a project. We then describe 4 key steps in developing a successful smart home innovation for people with mental and physical health conditions. These include (1) setting up the digital infrastructure, (2) ensuring the components of the system communicate, (3) ensuring that the system is designed for the intended population, and (4) engaging stakeholders. Recommendations on how to approach each of these steps are provided along with suggested literature that addresses additional considerations, guidelines, and equipment selection in more depth.

(JMIR Mhealth Uhealth 2022;10(4):e25116) doi:10.2196/25116

KEYWORDS

smart home; smart technology; mental health; physical health, eHealth; comorbidity; innovation; communication; connection; uHealth; ubiquitous health; digital health

Introduction

Smart home technologies present an unprecedented opportunity to improve health and health care. Enabled by the accessibility of wireless networks and computing devices inserted into everyday objects to create the internet of things [1], smart home technologies allow health practitioners to capture and monitor real-time data where behavioral anomalies such as irregular sleep patterns, inactivity, high heart rate levels, and sudden increase/decrease of weight could be indicative of psychological distress or the early stages of a health crisis. Within vulnerable populations, instrumental activities of daily living such as handling medication and self-care routines can start to decline rapidly [2,3]. By enabling health care personnel to remotely monitor physiological and mood changes in real time and by supporting communication between patients and providers,

smart home technologies can mitigate this decline by allowing individuals to manage their symptoms at home instead of waiting to visit health care facilities.

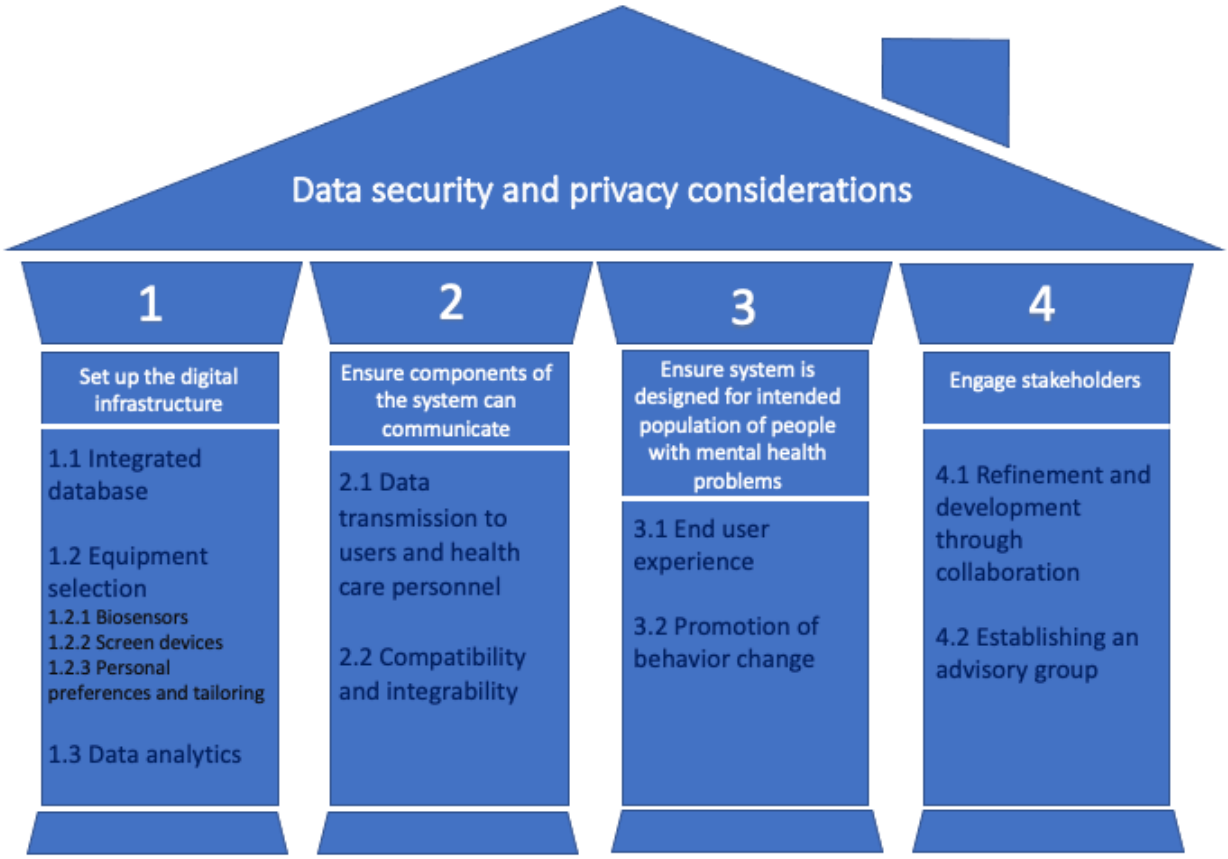
Health care systems are strained by the demand for mental and physical health care. Despite unprecedented federal investment in mental health and addiction services [4], over 2.3 million Canadians report unmet or partially met health care needs [5]. Demand for physical and mental health care has been exacerbated by the COVID-19 pandemic, even as it has increased barriers to access [6,7]. Smart home technologies can potentially reduce this demand by enabling users to monitor their health and live healthier lifestyles [8], particularly because physical activity has significantly declined in response to public health restrictions [9]. Reframing the health care system to deliver better coordinated, accessible care requires not only increased funds but also embracing innovative models of care delivery such as those afforded by smart home systems.

Beyond direct benefits to patients and the health care system, smart homes offer potentially significant wider economic benefits. While there may be upfront costs in purchasing, developing, and integrating the technology, there is potential for the innovation to be cost effective in the long term through a reduction in usage and resources for services. Researchers have reported a 38% decrease in administration costs pertaining to billing for care provisioning and care rescheduling within 8 years of introducing smart home care platforms [10]. There is the potential that smart technologies in the home would be helpful for individuals who require intensive support and experience barriers to, or disparities in, community care such as geographical location and inadequate travel arrangements, discrimination and stigma, and lack of resources [11].

Although technological studies investigating the use of apps [12], smartphones [13], and robotics [14] for mental health care exist, as well as for physical conditions such as Parkinson disease [15] and diabetes [16], research that incorporates aspects of both mental and physical health in one study remains limited, as is guidance for developing and implementing smart home–based interventions intended to support both mental and physical health. The contribution of this paper is to provide recommendations and considerations for future research projects looking to develop a smart technology innovation for mental and physical health care.

Reflecting on our own experience in this area [17], we have identified 4 key steps for developing a successful smart home–based intervention for supporting mental and physical health and health care that are outlined in Figure 1: (1) set up the digital infrastructure; (2) ensure components of the system can communicate; (3) ensure that the system is designed for the intended population; and (4) engage stakeholders. These 4 key steps are embedded within the overarching principle of data security and privacy. In the following, we describe each of the steps in detail and how they fit within this overarching principle, with the goal of supporting future research and implementation of smart home technologies for health.

Figure 1. Steps and considerations for a smart home technology innovation.



Data Security and Privacy Considerations

Crucial to any smart home innovation is the issue of data security and privacy by design. There are many challenges to maintaining anonymity and privacy. Even after data have been deidentified, there is still a risk of data linking the health record to the individual (eg, through a combination of location and rare attribute values). Recent research into data anonymized using Health Canada regulations has reported death records and social media to be the most likely sources of reidentification [18]. Furthermore, there are legal requirements in different jurisdictions that must be attended to in any project that collects data through smart devices.

In Canada, the provincial government of Ontario enacted the Personal Health Information Protection Act (PHIPA), a law that aims to maintain the security and privacy of personal health information of Ontarians by ensuring protection against unauthorized usage, theft, loss, or disclosure. Under this Act, agencies and organizations can send and receive deidentified health data, provided it is for the purposes of health system delivery, design, or planning. PHIPA is not only applicable to health information custodians, but also to custodians who have received health data, including health care providers and researchers. To ensure these standards are met, PHIPA asserts 10 fundamental principles [19]: (1) accountability; (2) identifying purposes; (3) knowledge and consent; (4) limiting collection; (5) limiting use, disclosure, and retention; (6) accuracy; (7) safeguards; (8) openness; (9) individual access; and (10) challenging compliance.

While PHIPA is specific to Ontario, there are many laws worldwide that hold similar principles [20,21]. It is important that these 10 principles are considered and followed when designing a smart technology innovation, both to ensure compliance with current governance and to enhance the efficacy, security, and privacy of data within a technology-based study. In the research environment, it is crucial that data be kept secure and participant identities remain anonymous. Further, in Ontario, the research team is required by PHIPA to fully investigate and report any form of breaches or losses of participant information to those participants identified in the data. In 2005, electronic health records require the consent of patients prior to their storage as outlined in the Personal Information Protection and Electronic Documents Act (PIPEDA) in Canada. This Act sets out to ensure that data are deidentified, and that personal information cannot be identifiable by a member of the public. Data linkages between organizations, whether they be private or public sector, are also governed by PIPEDA. However, this differs from PHIPA in that PIPEDA covers all data, whereas PHIPA focuses on health information data within Ontario. In the United States, the US Health Insurance Portability and Accountability Act (HIPAA) provides a set of guidelines for personal health data [22].

It should be highlighted that data stored on clouds, defined as a computing model that enables access to be shared or saved data regardless of the user's location or device used [23], may be subject to government inspection or surveillance depending on where the cloud servers are located. Under the Patriot Act (also known as the Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act of 2001), US law enforcement can issue a FISA (Foreign Intelligence Surveillance Act) Order or National Security Letter to any company that has a US office, US-based server, or conducts business within the United States, even if the data under investigation are stored on a server outside of the United States [24]. Similarly, the CLOUD (Clarifying Lawful Overseas Use of Data) Act allows US law enforcement to access the private data of American citizens stored on foreign servers [25]. In China, the passing of the Cybersecurity Law in November 2016, which legislated further cyberspace governance of data stored locally on servers based within China, has also led to tension and suspicion of government interference [26]. It is therefore important for potential studies to be aware of sensitive data being inspected by government agencies. It is recommended that researchers be aware of the data they are collecting and if they wish to maintain privacy, and that equipment and services are carefully selected.

A smart home innovation can be at risk of privacy breach depending on the security of the chosen wireless connectivity, for example, Wi-Fi, 3G, 4G, LTE (Long-Term Evolution), radio frequency identification, or wireless sensor networks [27] Wireless sensors and devices can be vulnerable to network-based attacks, resulting in exposure of personal information or malfunction. End-to-end encryption of all transmitted data and health records through tried and tested security protocols such as single sockets layer (SSL) can ensure the integrity and privacy of information and device operation. Additionally, any data integration across platforms must be accomplished without compromising privacy and integrity [28].

Finally, it is worth considering that there are the interrelationships between users within a home and the privacy boundaries that can arise when sharing devices among users or when technologies are situated in a shared space. An ethnographic study reported that one way of maintaining boundaries with technology within the smart home environment is to keep some rooms "technology free" [29]; this may be a valuable approach in certain shared-use situations. Individuals should be given the opportunity to share data with caregivers and members of their care team as well as be able to maintain privacy, both virtually and within the lived environment.

Having established the overarching principles of data security and privacy by design, we will next present our 4 steps for developing a successful smart home innovation.

Steps for Developing a Successful Smart Home Innovation

Set Up the Digital Infrastructure

Integrated Database

Simply monitoring the data coming from a multitude of devices and the devices' respective clouds or apps is not ideal for a health care professional, because it provides a disjointed view of a client's information. A centralized, integrated database capable of funneling the various forms of real-time data into

one accessible location can provide a more efficient approach to monitoring and tracking participant health data. Jensen [30] highlighted 5 key principles in designing a centralized integrated database: ease of use and implementation, built upon existing databases, expand installed base by persuasive tactics to gain momentum, making compatibility as simple as possible, and modularizing information infrastructures (ie, changes made do not affect the entire infrastructure).

Moreover, a centralized database can have other benefits such as providing a baseline rate of technology use and user activities whereby change can be observed and measured over time. From a research and a care perspective, this would provide an empirical basis for observation of mental and physical health progress/decline as well as precursors to health crises.

Security and privacy of a centralized integrated database can elicit concerns from health care providers and participants. This was evident in Iceland when 11% of the population opted out of a national database, causing the initiative to be scrapped [31]. This highlights the importance of ensuring that individual autonomy is respected and participants are well informed about the potential risks and benefits of their participation to obtain an informed consent. The consent process can be used to allay any fears of data mismanagement and provide assurances of data security. The participants need to understand how their data will be used and whether it will be for primary or secondary analyses. It is important to ensure that only authorized personnel, such as the participant's care provider, can access the health data and that this is clearly explained to both participants and care providers.

Outside of data security, there may also be concerns of a power failure or server crash that would compromise live data or result in a lack of access for care providers. It is recommended that administrators and technicians for an integrated database consider combining environmental and network redundancy with a robust virtual server environment to enable continuous operations with minimal or no impact on usage or data transmission. This can also be utilized during peak overloads or during maintenance to ensure functionality.

Equipment Selection

Overview

In the following, we discuss the selection of biosensors and screen devices, and we discuss the issue of tailoring devices to the needs of populations. In our experience, for all of these types of devices, it is advantageous to choose screen devices and biosensors that are easily and commercially available (ie, able to be purchased in shops/online) as opposed to expensive clinical equipment that are harder to purchase or replace.

Biosensors—Weigh Scales, Blood Pressure Monitors, Glucometers, and Wearable Activity Trackers

The interplay of physical and mental health conditions often complicate help-seeking and adherence to treatment, which worsens the prognosis of all present diseases [32,33]. The prevalence of chronic physical conditions, such as diabetes and cardiovascular disease, are higher among people with mental illness [34,35]. To account for the rise in comorbidities, a comprehensive smart home intervention equipped with noninvasive and nonintrusive sensors (wearable activity trackers, glucometers, and weigh scales) can be viable diagnostic tools for health care professionals [36,37]. The embedded sensors within these devices capture important physiological parameters such as cardiorespiratory function, blood sugar levels, weight/BMI, gait analysis, and sleep patterns that are then funneled into a centralized database, capable of logging and analyzing the collected data for anomalies.

Wearable sensor-based health monitoring systems should achieve maximal usability and reduce operator discomfort. Sensors that are embedded into textile fabrics or surface mounted directly to the body require a stable sensor-skin interface to ensure high signal accuracy and durability [36]. From an ergonomic standpoint, the selected hardware should be comfortable, flexible, small, and unobtrusive when attached to the body. It should also ensure minimal risk of harm to users and compliant with industry standards (ie, nontoxic, nonreactive, and manufactured from hypoallergic materials).

These devices should have low power requirements and should exhibit "always-on" connectivity as frequent removal for charging or syncing prevents seamless integration with daily activity and presents a barrier to long-term utilization. Tech-enabled, noncritical monitoring used in tandem with standard health services has potential not only to alleviate the workload on health care providers, but also can be used to inform accurate symptom reports, diagnoses, and prompt referrals to specialized care.

Screen Devices—Smartphones, Tablets, and Monitors; Apps

Connectivity with care providers and access to resources have become increasingly enhanced with modern hardware and software. Research has reported that usage can be measured (ie, number of SMS text messages sent, physical movement) to predict mood disorder changes [38]. The use of screen devices in combination with specially designed apps can render positive outcomes. A systematic review on mental health apps used on smartphones and tablets revealed significant reductions in depression, stress, and substance use [39]. Another review of physical health and self-care apps revealed an increase in physical activity, weight management, smoking cessation, and medication adherence with favorable feasibility and usability [40]. We recommend that smart home technology interventions include screen devices, such as tablets, monitors, or smartphones. Screen devices should be equipped with secure apps to support remote telemonitoring and provide access to personal health information. They should also provide secure communication between participants and their health care providers (including electronic messaging and videoconferencing) and provide the ability to customize and display screen prompts including real-time notifications to facilitate self-care.

Consequently, health care providers can respond swiftly to significant deviations in the patient's normal vital signs by notifying the patient through the screen devices and scheduling ad hoc virtual care. Communication through the screen devices should be bidirectional so patients can not only acknowledge

and respond to notifications from their health care providers, but can also request additional help or support, if required.

Screen devices should also include features that support mood tracking, such as standard or client-specific self-assessments. Not only does mood tracking reduce the logistic burden of data collection, but it also has the potential to enrich clinical practice by offering novel methods of monitoring psychopathology [41]. Conventional, cross-sectional surveys and clinical interactions have limited impact because they collect retrospective rather than real-time data. Mobile apps with mood-tracking features represent a feasible method for capturing "in the moment" patient data through experience sampling methods [42]. Changes in mood could also be reflective of other physical health symptomology that may be present or alleviated. This provides a more detailed understanding of mental health phenomenology because daily fluctuations in mood are linked to time-stamped activities and social contexts, thereby elucidating the dynamic relationship between environment and symptomatic experiences [41-43]. A technological innovation using mood tracking technology in combination with biosensors can further outline the bidirectional effects of mental and physical health.

Consideration of Personal Preferences and Tailoring

Models of intervention delivery leveraging the ubiquity of technological access must be tailored to the unique challenges faced by people with mental and physical health problems in order to be successful. To be most impactful, components of the intervention should be client tailored and thus, developed and integrated under a framework of validated health behavior theories [44]. In the Lawson Health Smart Home study [17], participants were able to add/remove equipment as and when was needed, or due to personal preference. For example, one participant originally declined the medication dispenser but then later requested it to assist with medication adherence. The dispenser was installed and the participant reported no missed doses thereafter.

It is very important to consider the potential negative consequences on mental health following the introduction of technology. For example, the use of smart mirrors (mirrors with touchscreen and connectivity capabilities) and voice-activated devices could exacerbate paranoia or delusions. Although user preference is paramount, clinical decision makers are needed for input into the appropriateness of devices. The introduction of technology should provide support and care, and smart technology interventions must be flexible in accounting for specific illnesses, symptoms, and comorbidities.

Data Analytics

Health apps can provide analytical data to users by monitoring their health and providing feedback. Dimitrov [45] noted that the use of artificial intelligence aims to identify and analyze patterns of data from one individual and then compare those with similar patterns to make predictive recommendations, thereby acting as a personal coach to the individual. An example of this would be number of steps or calories burned, and how that aligns with the individual's goals and targets. It could be further argued that appropriate application of artificial intelligence would help triage patients by empowering participants and lessening the burden on health care services and clinicians [45]. Data derived from health devices can be combined with existing data sets to provide greater contextual understandings. Although analytic capabilities are useful, it is crucial to remember that population and individual metrics can vary differently and can be determined through differing methodologies/models. For example, blood pressure monitoring would not be generalizable to the population due to measurement errors based on the time of day [46]. Taking individual factors into account would be a key recommendation when developing an algorithm or a predictive model as errors can occur and erroneous assumptions could be made.

It is recommended that future research studies follow these 6 best practice steps [47] for analyzing health data from wearable devices and health apps:

1. set a robust research question where access to the necessary data is available, including the data on the device and publicly available data sets for comparisons (eg, life expectancy, prevalence rates);
2. prepare data for analysis to anticipate inaccuracies and missing data as well as removing outliers and erroneous values, ascertaining sensor accuracy, and performing trial runs prior to implementation;
3. verify the data set by comparing with current literature and gold-standard data sets to ensure consistency or find inconsistencies;
4. analyze data that answer the research question(s), establishing causal relationships and the possibility of needing propensity scoring models to minimize confounding variables;
5. check the robustness of conclusions and establish the validity of the findings;
6. knowledge dissemination including sharing data sets and coding while maintaining privacy and governance standards.

In the Lawson Health Smart Home innovation, questionnaires were programmed so that if a participant expressed suicidal thoughts or ideations, a notification was immediately sent by SMS text message or email to their care provider(s). It is important for any smart technology study that trends can be observed and saved for future analyses and to provide a bigger picture of the participant's current health with contingencies in the event there are concerning data trends.

Ensure Components of the System Can Communicate

Data Transmission to Users and Health Care Personnel

Unprocessed data captured by noninvasive sensors are transmitted to a processing node using a low-power and short-range communication protocol such as Bluetooth, ZigBee, ANT, or near-field communications (NFC) [36,48-53]. The processing node is an advanced platform that first collects and filters the data, and then executes advanced analysis and decision algorithms. Common examples include smartphones, tablets, computers, and personal digital assistants. The processing node functions as a gateway, or a central hub, that receives raw data input from wearables and transmits measured data output to a secure server located in a remote health care facility. Depending on what processing node is used, results may be stored,

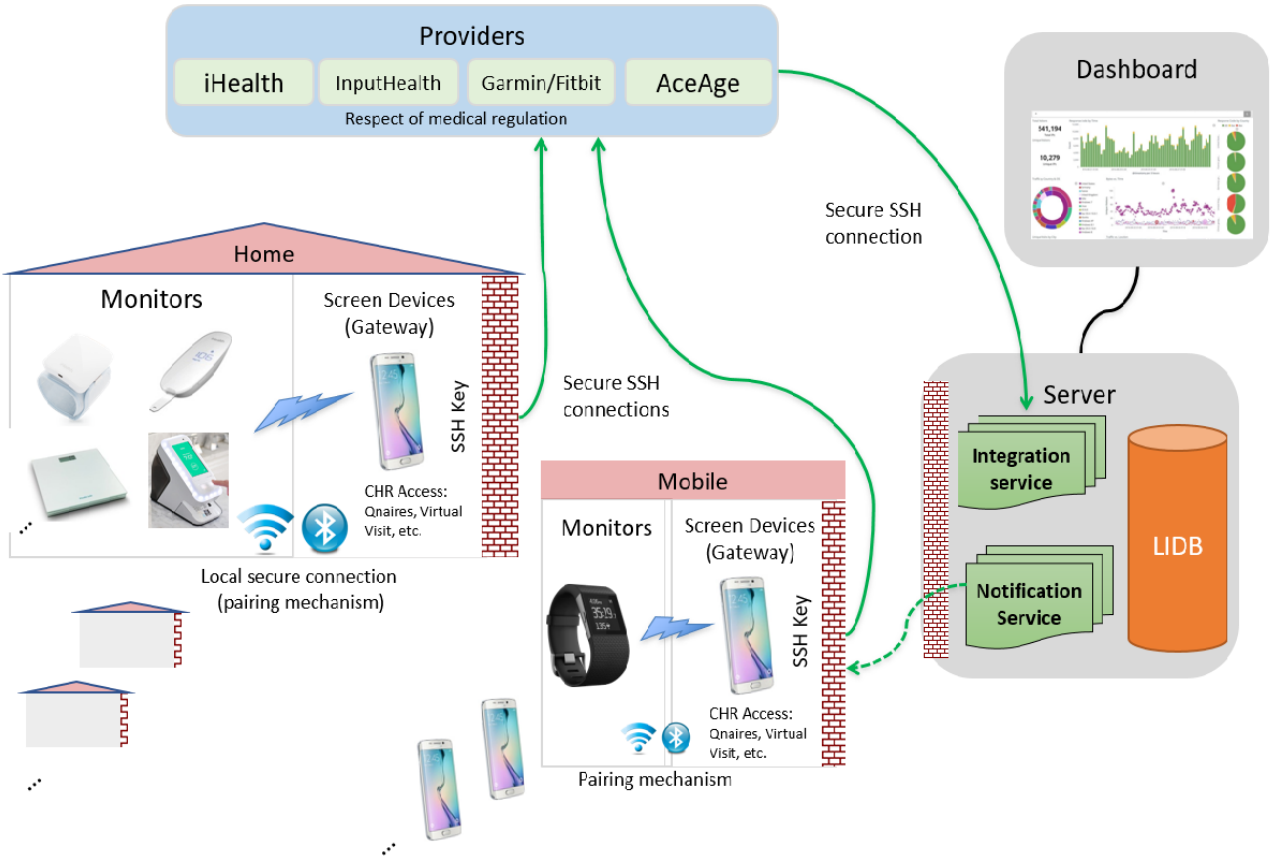
translated, and displayed on a user interface. Interventions consisting of only a few sensors can send data directly to the processing node, while more complex systems with several sensor units can first gather data through a central body area network [36]. Processed data transmitted to the hospital server over the internet requires a long-range communication protocol.

To ensure data transmission, the processing node should achieve reliable, stable data transmission without any interference and present a low risk of being hacked. In a multisensor body area network system, the central processing node requires a significant amount of power to handle the large influx of data; therefore, power consumption levels must be lowered to sustain long-term utilization [36]. Likewise, the transmission of measured data outputted from the processing node to a centralized integrated database through the internet will require appropriate encryption (ie, public key infrastructure, SSL), authorization, and authentication algorithms to ensure personal health information is secure and protected [36,54,55].

Compatibility and Integrability

Compatibility is a key issue when connecting multiple devices for a smart technology innovation. A major aspect of building a smart technology innovation is for the individual pieces of equipment to be able to communicate with one another, or that they are all compatible with 1 central data center. See Figure 2 for an example of the smart home system from the Lawson Health study [17]. With multiple operating systems currently available (eg, Android, iOS, Windows, Linux, Raspberry), it is crucial that the technologies are capable of interacting with one another and able to transmit data seamlessly without any drop in performance. To implement a centralized integration of data originating from different sources, it is essential to consider software formats so that the end users can be provided with a translated and unified view of the accumulated data. It may require the data extraction and synchronization processes, considering data might be coming from myriad sources using different transmission schedules and rates. This is where implementation of processes such as ETL (extract, transform, load) is advocated. Integrability is a key component of developing a smart home innovation as individual devices need to be integrated seamlessly into the individual's care plan as well as meeting the purposes of the system. Furthermore, a smart home innovation could result in the integration of multiple members of the care team or additional service providers consulting on all of the data collected as opposed to their own individualized data.

Figure 2. System architecture for the Lawson Smart Homes Innovation [17]. CHR: collaborative health record; LIDB: Lawson Integrated Database; Qnaires: Questionnaires from CHR; SSH: secure shell.



Ensure That the System Is Designed for the Intended Population

End User Experience

Smart technology interventions for people with mental and physical health problems should be designed with the goal of optimizing the user experience. It is imperative that interactions with the selected equipment are as simple and efficient as possible. A smart home innovation should connect devices to anticipate user needs, automate complex configurations, and be "smart" enough to make independent, intelligent decisions. The limited space for micro interactions on wearable devices should be intuitively optimized so that features can be easily navigated on the main menu [56]. In addition, the innovation should have a micro feedback display that provides immediate, quantitative feedback on biometrics related to the users' exercise quality, cardiorespiratory function, and sleep patterns.

On the hub, the data are processed and translated into usable information and the results are displayed on a mobile app that synchronizes to the wearable hardware. The mobile app is the component with which users most frequently interact and should connect with the entire suite of selected wearable devices. The app should be able to chart their progress with data visualization tools such as calendars and graphs and provide relevant information (eg, steps, distance, weight, sleep cycle) that is easily accessible and interpreted. The app interface should use colors, animations, and typography congruent with the intended purpose of a mobile app and embrace minimal cognitive overload principles (ie, avoid too much content, irrelevant steps, or inconsistent formatting).

Promotion of Behavior Change

In addition to presenting the data, devices should include functionality that promotes sustained behavior change where appropriate. Based on principles of positive reinforcement, strategies to engage users should address all 3 components of the habit formation loop [56]: cues (push notifications, alarms, vibrations), behaviors (prompts to exercise, sleep, hydrate), and rewards (badges, level progressions, gamification). Research has shown that sharing personal information with friends and family increases the likelihood of attaining goals [57,58]. Mobile apps selected or developed for smart home interventions should support additional aspects of care provisions such as secure virtual care (messaging, audio, and visual), mood tracking features, and active prompts for medication, activity, or appointment reminders.

Engage Stakeholders

Refinement and Development Through Collaboration

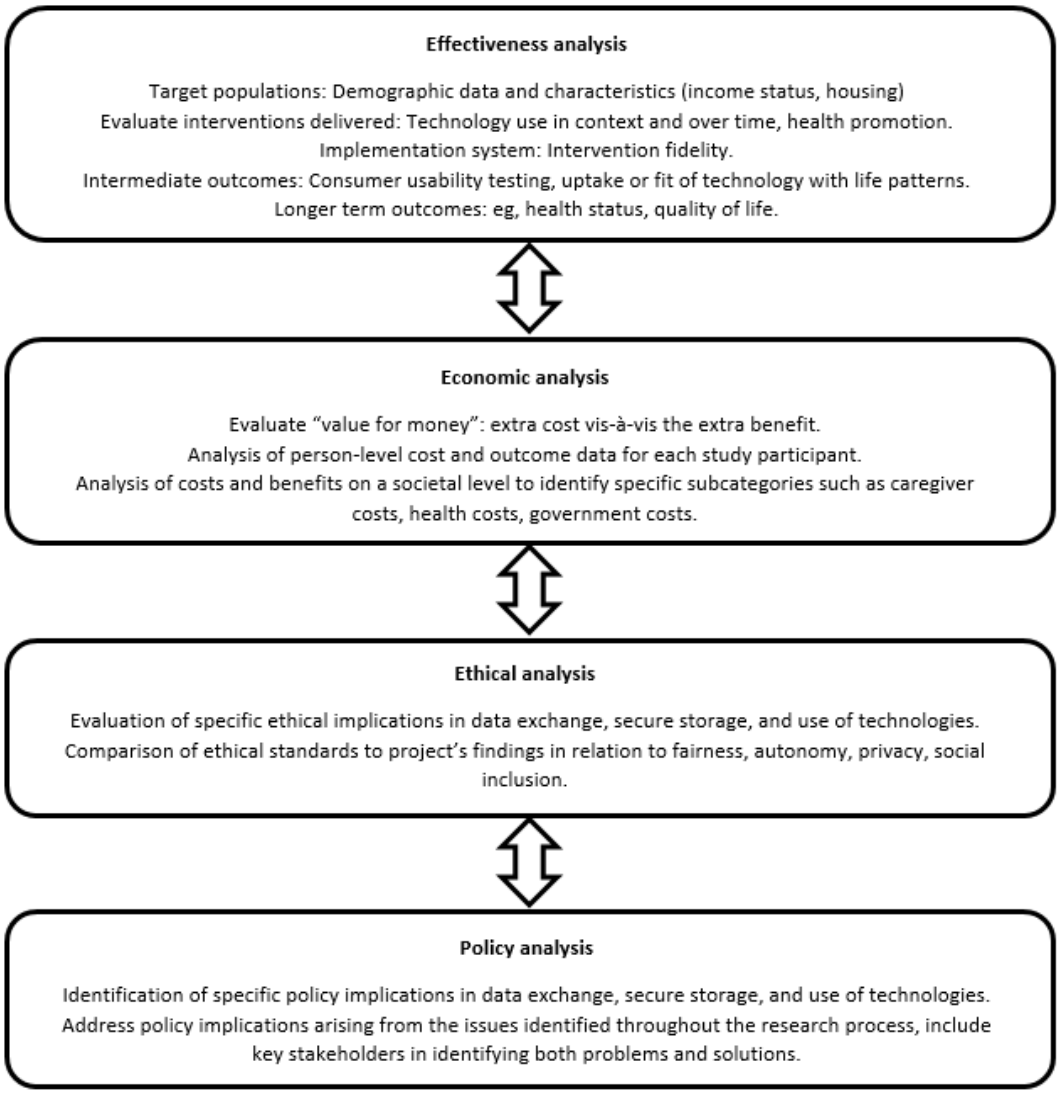
Consultation with stakeholders is a crucial component of developing a smart technology innovation. It is important to identify from the outset as to who would have an interest in this innovation and whether they can assist in its development, refinement, and implementation. This is particularly of great value when the new innovation is untested, unproven, or may contain risks such as privacy. Hart and Sharma [59] note that many companies tend to focus on stakeholders that are known to them or hold powerful positions but fail to engage with peripheral stakeholders. This then leads to a lack of competitive imagination and positive disruptive change. It was suggested by Wright [60] that engaging stakeholders prior to deployment of new information technology can help to mitigate ethical implications by ensuring measures are taken and risks adequately examined. This is also true of research: by embracing and encouraging ideas from stakeholders outside of the research environment, the project can develop a greater depth of ideas and imagination. Roles in provision, oversight, and management of the intervention should involve an interdisciplinary team of health care providers and experts of the mental and physical illnesses the intervention is seeking to address and support.

Establishing an Advisory Group

It is recommended that stakeholders be encouraged to provide input throughout the design process through the formation of an advisory group. A co-design approach through the collaborative efforts of an advisory group can allow for reductions in oversights, greater suggestions for improvement, and ensuring targets are met. The objectives of an advisory group are to ensure the project remains on track, help the project overcome obstacles that may arise, and help align the technology with present and future opportunities for scaling up or further development.

This group ideally consists of end users, consumers, service providers, policy decision makers, industry representatives, analysts, and researchers. End users and consumers should be consulted throughout development which can involve "citizen juries" [61], focus groups, and pilot testing. As well, analysts and researchers within the advisory group should contribute to the evaluation of the intervention. The evaluation depends on the research questions and hypotheses of the project. A standardized evaluation framework for smart technology intervention that covers a broad range of implications would be recommended [62] (Figure 3). It is important that not just the effectiveness of the intervention is measured, but also inferences for sustainability beyond the study; cost is compared with usual care; and ethical considerations are obtained. In summary, a comprehensive approach to evaluation and refinement of a smart home intervention from all interested parties allows for risk management, improved quality control, and enhanced innovation.

Figure 3. Standardized evaluation framework for smart technology mental health interventions [62].



Conclusion

The overarching principle of data security and privacy together with our 4 steps were identified from our experience developing a smart home intervention for a target population that included end users who were particularly struggling with physical and mental health conditions. However, they are equally applicable to other interventions that address many different mental and physical health concerns. Going forward, we see huge potential for smart home research and implementation work that takes a holistic view of the end user. Smart home technologies have the potential to support all aspects of health by targeting not only health care issues but also by facilitating more social inclusion and better health behaviors, leading to improved quality of life.

Conflicts of Interest

None declared.

References

1. Wortmann F, Flüchter K. Internet of Things. Bus Inf Syst Eng 2015 Mar 27;57(3):221-224. [doi: 10.1007/s12599-015-0383-3]
2. Kazama M, Kondo N, Suzuki K, Minai J, Imai H, Yamagata Z. Early impact of depression symptoms on the decline in activities of daily living among older Japanese: Y-HALE cohort study. Environ Health Prev Med 2011 Nov 03;16(3):196-201. [doi: 10.1007/s12199-010-0186-6] [Medline: 21431794]
3. De Vriendt P, Gorus E, Cornelis E, Velghe A, Petrovic M, Mets T. The process of decline in advanced activities of daily living: a qualitative explorative study in mild cognitive impairment. Int. Psychogeriatr 2012 Feb 03;24(6):974-986. [doi: 10.1017/s1041610211002766] [Medline: 22301014]

4. Bartram M. Making the most of the federal investment of $5 billion for mental health. CMAJ 2017 Nov 06;189(44):E1360-E1363. [doi: 10.1503/cmaj.170738] [Medline: 29109209]
5. Statistics Canada. Mental Health Care Needs 2018. Statistics Canada. 2019 Oct 07. URL: https://www.150.statcan.gc.ca/n1/pub/82-625-x/2019001/article/00011-eng.htm [accessed 2020-04-22]
6. Hoffman D. Increasing access to care: Telehealth during COVID-19. J Law Biosci 2020 Jun 16;7(1):1-15. [doi: 10.1093/jlb/lsaa043] [Medline: 32843985]
7. Kannarkat JT, Smith NN, McLeod-Bryant SA. Mobilization of telepsychiatry in response to COVID-19 - moving toward 21st century access to care. Adm Policy Ment Health 2020 Apr 24;47(4):489-491. [doi: 10.1007/s10488-020-01044-z] [Medline: 32333227]
8. Buoite Stella A, Ajčević M, Furlanis G, Cillotto T, Menichelli A, Accardo A, et al. Smart technology for physical activity and health assessment during COVID-19 lockdown. J Sports Med Phys Fitness 2021 Mar;61(3):452-460. [doi: 10.23736/s0022-4707.20.11373-2] [Medline: 33092330]
9. Stockwell S, Trott M, Tully M, Shin J, Barnett Y, Butler L, et al. Changes in physical activity and sedentary behaviours from before to during the COVID-19 pandemic lockdown: a systematic review. BMJ Open Sport Exerc Med 2021 Feb 01;7(1):e000960. [doi: 10.1136/bmjsem-2020-000960] [Medline: 34192010]
10. Vannieuwenborg F, Van der Auwermeulen T, Van Ooteghem J, Jacobs A, Verbrugge S, Colle D. Evaluating the economic impact of smart care platforms: qualitative and quantitative results of a case study. JMIR Med Inform 2016 Oct 31;4(4):e33 [FREE Full text] [doi: 10.2196/medinform.5012] [Medline: 27799137]
11. Gibbons MC. A historical overview of health disparities and the potential of eHealth solutions. J Med Internet Res 2005 Oct 04;7(5):e50 [FREE Full text] [doi: 10.2196/jmir.7.5.e50] [Medline: 16403714]
12. Hollis C, Morriss R, Martin J, Amani S, Cotton R, Denis M, et al. Technological innovations in mental healthcare: harnessing the digital revolution. Br J Psychiatry 2015 Apr;206(4):263-265. [doi: 10.1192/bjp.bp.113.142612] [Medline: 25833865]
13. Clough BA, Casey LM. The smart therapist: A look to the future of smartphones and mHealth technologies in psychotherapy. Professional Psychology: Research and Practice 2015 Jun;46(3):147-153. [doi: 10.1037/pro0000011]
14. Riek L. Robotics technology in mental health care. In: Luxton DD, editor. Artificial Intelligence in Behavioral and Mental Health Care. San Diego, CA: Elsevier Academic Press; 2016:185-203.
15. Pasluosta CF, Gassner H, Winkler J, Klucken J, Eskofier BM. An emerging era in the management of Parkinson's disease: wearable technologies and the Internet of Things. IEEE J. Biomed. Health Inform 2015 Nov;19(6):1873-1881. [doi: 10.1109/jbhi.2015.2461555] [Medline: 26241979]
16. Sangave N, Aungst T, Patel D. Smart connected insulin pens, caps, and attachments: a review of the future of diabetes technology. Diabetes Spectrum 2019 Nov;32(4):378-384. [doi: 10.2337/ds18-0069] [Medline: 31798296]
17. Forchuk C, Serrato J, Rudnick A, Corring D, Mann R, Frampton B. An interconnected smart technology system for individuals with mental illness living in the community transitional hospital apartments. In: Pagán J, Mokhtari M, Aloulou H, Abdulrazak B, Cabrera M, editors. How AI Impacts Urban Living and Public Health. ICOST 2019. Lecture Notes in Computer Science. Cham, Switzerland: Springer; Oct 07, 2019:131-142.
18. Branson J, Good N, Chen J, Monge W, Probst C, El Emam K. Evaluating the re-identification risk of a clinical study report anonymized under EMA Policy 0070 and Health Canada Regulations. Trials 2020 Feb 18;21(1):A200. [doi: 10.1186/s13063-020-4120-y] [Medline: 32070405]
19. eHealth Ontario. Personal Health Information Privacy Policy. eHealth Ontario. 2016. URL: https://www.ehealthontario.on.ca/images/uploads/pages/documents/PHI_PrivacyPolicy_EN.pdf [accessed 2020-06-09]
20. Abouelmehdi K, Beni-Hessane A, Khaloufi H. Big healthcare data: preserving security and privacy. J Big Data 2018 Jan 9;5:A1. [doi: 10.1186/s40537-017-0110-7]
21. Committee on the Role of Institutional Review Boards in Health Services Research Data Privacy Protection. International comparisons of data privacy standards. In: Protecting Data Privacy in Health Services Research. Washington, DC: National Academy Press; 2000:148-150.
22. US Department of Health and Human Services. Health Insurance Portability and Accountability Act of 1996. US Department of Health and Human Services. 1996. URL: https://aspe.hhs.gov/report/health-insurance-portability-and-accountability-act-1996 [accessed 2020-03-10]
23. Jadeji Y, Modi K. Cloud Computing - Concepts, Architecture and Challenges. 2012 Presented at: International Conference on Computing, Electronics and Electrical Technologies (ICCEET); March 21/22, 2012; Kumaracoil, India p. 21 URL: https://ieeexplore.ieee.org/abstract/document/6203873 [doi: 10.1109/icceet.2012.6203873]
24. Lakatos A. The USA Patriot Act and the Privacy of Data Stored in the Cloud. Mayer Brown. 2012 Jan 18. URL: https://www.mayerbrown.com/en/perspectives-events/publications/2012/01/the-usa-patriot-act-and-the-privacy-of-data-stored [accessed 2020-03-10]
25. US Department of Justice. Promoting Public Safety, Privacy, and the Rule of Law Around the World: The Purpose and Impact of the CLOUD Act. US Department of Justice. 2019 Apr 05. URL: https://www.justice.gov/opa/press-release/file/1153446/download [accessed 2020-07-20]
26. Qi A, Shao G, Zheng W. Assessing China's cybersecurity law. Computer Law & Security Review 2018 Dec;34(6):1342-1354. [doi: 10.1016/j.clsr.2018.08.007]

27. Park M, Oh H, Lee K. Security risk measurement for information leakage in IoT-based smart homes from a situational awareness perspective. Sensors 2019 May 09;19(9):2148. [doi: 10.3390/s19092148] [Medline: 31075883]
28. Zhang R, Liu L. Security Models Requirements for Healthcare Application Clouds. New York, NY: IEEE; 2010 Aug 26 Presented at: IEEE 3rd International Conference on Cloud Computing; July 5-10, 2010; Miami, FL p. 268-275. [doi: 10.1109/cloud.2010.62]
29. Burrows A, Coyle D, Gooberman-Hill R. Privacy, boundaries and smart homes for health: An ethnographic study. Health & Place 2018 Mar;50:112-118. [doi: 10.1016/j.healthplace.2018.01.006] [Medline: 29428688]
30. Jensen TB. Design principles for achieving integrated healthcare information systems. Health Informatics J 2013 Mar 13;19(1):29-45. [doi: 10.1177/1460458212448890] [Medline: 23486824]
31. Anderson R. Confidentiality and connecting for health. Br J Gen Prac 2008 Feb;58(547):75-76 [FREE Full text] [doi: 10.3399/bjgp08X263992] [Medline: 18307846]
32. Sartorius N. Comorbidity of mental and physical disorders: a main challenge for medicine of the 21st century. Shanghai Arch Psychiatry 2013 Apr;25(2):68-69. [doi: 10.3969/j.issn.1002-0829.2013.02.002] [Medline: 24991137]
33. Prince M, Patel V, Saxena S, Maj M, Maselko J, Phillips MR, et al. No health without mental health. The Lancet 2007 Sep 04;370(9590):859-877. [doi: 10.1016/s0140-6736(07)61238-0] [Medline: 17804063]
34. Huang K, Su T, Chen T, Chou Y, Bai Y. Comorbidity of cardiovascular diseases with mood and anxiety disorder: a population based 4-year study. Psychiatry Clin Neurosci 2009 Jun;63(3):401-409. [doi: 10.1111/j.1440-1819.2009.01974.x] [Medline: 19566773]
35. Katon WJ. The comorbidity of diabetes mellitus and depression. Am J Med 2008 Nov;121(11 Suppl 2):S8-S15. [doi: 10.1016/j.amjmed.2008.09.008] [Medline: 18954592]
36. Majumder S, Mondal T, Deen M. Wearable sensors for remote health monitoring. Sensors 2017 Jan 12;17(1):130. [doi: 10.3390/s17010130] [Medline: 28085085]
37. Deen MJ. Information and communications technologies for elderly ubiquitous healthcare in a smart home. Pers Ubiquit Comput 2015 Jun 26;19(3-4):573-599. [doi: 10.1007/s00779-015-0856-x]
38. Beiwinkel T, Kindermann S, Maier A, Kerl C, Moock J, Barbian G, et al. Using smartphones to monitor bipolar disorder symptoms: a pilot study. JMIR Ment Health 2016 Jan 06;3(1):e2 [FREE Full text] [doi: 10.2196/mental.4560] [Medline: 26740354]
39. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
40. Rathbone AL, Prescott J. The use of mobile apps and SMS messaging as physical and mental health interventions: systematic review. J Med Internet Res 2017 Aug 24;19(8):e295 [FREE Full text] [doi: 10.2196/jmir.7740] [Medline: 28838887]
41. van Os J, Verhagen S, Marsman A, Peeters F, Bak M, Marcelis M, et al. The experience sampling method as an mHealth tool to support self-monitoring, self-insight, and personalized health care in clinical practice. Depress Anxiety 2017 May 23;34(6):481-493. [doi: 10.1002/da.22647] [Medline: 28544391]
42. Torous J, Friedman R, Keshavan M. Smartphone ownership and interest in mobile applications to monitor symptoms of mental health conditions. JMIR Mhealth Uhealth 2014 Jan 21;2(1):e2 [FREE Full text] [doi: 10.2196/mhealth.2994] [Medline: 25098314]
43. Kimhy D, Delespaul P, Ahn H, Cai S, Shikhman M, Lieberman JA, et al. Concurrent measurement of "Real-World" stress and arousal in individuals with psychosis: assessing the feasibility and validity of a novel methodology. Schizophr Bull 2010 Nov;36(6):1131-1139. [doi: 10.1093/schbul/sbp028] [Medline: 19429846]
44. Naslund JA, Aschbrenner KA, Kim SJ, McHugo GJ, Unützer J, Bartels SJ, et al. Health behavior models for informing digital technology interventions for individuals with mental illness. Psychiatr Rehabil J 2017 Sep;40(3):325-335. [doi: 10.1037/prj0000246] [Medline: 28182469]
45. Dimitrov DV. Medical Internet of Things and big data in healthcare. Healthc Inform Res 2016 Jul;22(3):156-163. [doi: 10.4258/hir.2016.22.3.156] [Medline: 27525156]
46. Li Y, Sperrin M, Belmonte M, Pate A, Ashcroft DM, van Staa TP. Do population-level risk prediction models that use routinely collected health data reliably predict individual risks? Sci Rep 2019 Aug 2;9(1):11222. [doi: 10.1038/s41598-019-47712-5] [Medline: 31375726]
47. Hicks JL, Althoff T, Sosic R, Kuhar P, Bostjancic B, King AC, et al. Best practices for analyzing large-scale health data from wearables and smartphone apps. NPJ Digit Med 2019 Jun 3;2:45. [doi: 10.1038/s41746-019-0121-1] [Medline: 31304391]
48. Dementyev A, Hodges S, Taylor S, Smith J. Power consumption analysis of Bluetooth low energy, ZigBee and ANT sensor nodes in a cyclic sleep scenario. 2013 Oct 03 Presented at: IEEE International Wireless Symposium (IWS); April 14-18, 2013; Beijing, China. [doi: 10.1109/ieee-iws.2013.6616827]
49. Suzuki T, Tanaka H, Minami S, Yamada H, Miyata T. Wearable wireless vital monitoring technology for smart health care. 2013 May 30 Presented at: 7th International Symposium on Medical Information and Communication Technology (ISMICT); March 6-8, 2013; Tokyo, Japan. [doi: 10.1109/ismict.2013.6521687]

50. Malhi K, Mukhopadhyay SC, Schnepper J, Haefke M, Ewald H. A Zigbee-based wearable physiological parameters monitoring system. IEEE Sensors J 2012 Mar;12(3):423-430. [doi: 10.1109/jsen.2010.2091719]
51. Valchinov E, Antoniou A, Rotas K, Pallikarakis N. Wearable ECG System for Health and Sports Monitoring. 2014 Dec 05 Presented at: 4th International Conference on Wireless Mobile Communication and Healthcare - Transforming Healthcare through Innovations in Mobile and Wireless Technologies?; November 3-5, 2014; Athens, Greece. [doi: 10.4108/icst.mobihealth.2014.257236]
52. Coskun V, Ozdenizci B, Ok K. A survey on Near Field Communication (NFC) technology. Wireless Pers Commun 2012 Dec 1;71(3):2259-2294. [doi: 10.1007/s11277-012-0935-5]
53. Pang Z, Zheng L, Tian J, Kao-Walter S, Dubrova E, Chen Q. Design of a terminal solution for integration of in-home health care devices and services towards the Internet-of-Things. Enterprise Information Systems 2015;9(1):86-116. [doi: 10.1080/17517575.2013.776118]
54. Corchado J, Bajo J, Abraham A. GerAmi: improving healthcare delivery in geriatric residences. IEEE Intell Syst 2008 Mar 21;23(2):19-25. [doi: 10.1109/mis.2008.27]
55. Stav E, Walderhaug S, Mikalsen M, Hanke S, Benc I. Development and evaluation of SOA-based AAL services in real-life environments: A case study and lessons learned. Int J Med Inform 2013 Nov;82(11):e269-e293. [doi: 10.1016/j.ijmedinf.2011.03.007] [Medline: 21481634]
56. Endeavour Partners. Inside Wearables Part 1: How behavior change unlocks long-term engagement. Medium. 2014 Jan. URL: https://medium.com/@endeavourprtnrs/inside-wearable-how-the-science-of-human-behavior-change-offers-the-secret-to-long-term-engagement-a15b3c7d4cf3 [accessed 2020-04-22]
57. Cialdini RB. Harnessing the science of persuasion. Harv Bus Rev 2001 Oct;79(9):72-79. [doi: 10.1038/scientificamerican0201-76]
58. Ellison N, Steinfield C, Lampe C. The benefits of Facebook "friends": Social capital and college students' use of online social network sites. J Comput Mediat Commun 2007 Jul;12(4):1143-1168. [doi: 10.1111/j.1083-6101.2007.00367.x]
59. Hart SL, Sharma S. Engaging fringe stakeholders for competitive imagination. Acad Manag Exec 2004 Feb;18(1):7-18. [doi: 10.5465/ame.2004.12691227]
60. Wright D. A framework for the ethical impact assessment of information technology. Ethics Inf Technol 2011 Sep;13(3):199-226. [doi: 10.1007/s10676-010-9242-6]
61. Menon D, Stafinski T. Engaging the public in priority-setting for health technology assessment: findings from a citizens' jury. Health Expect 2008 Sep;11(3):282-293. [doi: 10.1111/j.1369-7625.2008.00501.x] [Medline: 18816324]
62. Forchuk C, Rudnick A, MacIntosh J, Bukair F, Hoch J. Evaluation framework for smart technology mental health interventions. In: Chang C, Chiari L, Cao Y, Jin H, Mokhtari M, Aloulou H, editors. Lect Notes Comp Sci. Cham, Switzerland: Springer; 2016:203-210.

Abbreviations

CLOUD: Clarifying Lawful Overseas Use of Data Act
ETL: extract, transform, load
FISA: Foreign Intelligence Surveillance Act
HIPAA: US Health Insurance Portability and Accountability Act
LTE: Long-Term Evolution
NFC: near field communications
PHIPA: Personal Health Information Protection Act
PIPEDA: Personal Information Protection and Electronic Documents Act
SSL: single sockets layer

Edited by L Buis, A Mavragani; submitted 23.10.20; peer-reviewed by T Ungar, S Chen; comments to author 15.12.20; revised version received 10.02.21; accepted 22.02.22; published 29.04.22.

Please cite as:
Forchuk C, Serrato J, Lizotte D, Mann R, Taylor G, Husni S
Developing a Smart Home Technology Innovation for People With Physical and Mental Health Problems: Considerations and Recommendations
JMIR Mhealth Uhealth 2022;10(4):e25116
URL: https://mhealth.jmir.org/2022/4/e25116
doi:10.2196/25116
PMID:35486422

Review

Deep Learning in mHealth for Cardiovascular Disease, Diabetes, and Cancer: Systematic Review

Andreas Triantafyllidis[1*], PhD; Haridimos Kondylakis[2*], PhD; Dimitrios Katehakis[2*], MS; Angelina Kouroubali[2*], PhD; Lefteris Koumakis[2*], PhD; Kostas Marias[2*], PhD; Anastasios Alexiadis[1*], PhD; Konstantinos Votis[1*], PhD; Dimitrios Tzovaras[1*], PhD

[1]Information Technologies Institute, Centre for Research and Technology Hellas, Thessaloniki, Greece

[2]Institute of Computer Science, Foundation for Research and Technology Hellas, Heraklion, Greece

[*]all authors contributed equally

Corresponding Author:
Andreas Triantafyllidis, PhD
Information Technologies Institute
Centre for Research and Technology Hellas
6th Km Charilaou-Thermi
Thessaloniki, 57001
Greece
Phone: 30 6946059255
Email: atriand@iti.gr

Abstract

Background: Major chronic diseases such as cardiovascular disease (CVD), diabetes, and cancer impose a significant burden on people and health care systems around the globe. Recently, deep learning (DL) has shown great potential for the development of intelligent mobile health (mHealth) interventions for chronic diseases that could revolutionize the delivery of health care anytime, anywhere.

Objective: The aim of this study is to present a systematic review of studies that have used DL based on mHealth data for the diagnosis, prognosis, management, and treatment of major chronic diseases and advance our understanding of the progress made in this rapidly developing field.

Methods: A search was conducted on the bibliographic databases Scopus and PubMed to identify papers with a focus on the deployment of DL algorithms that used data captured from mobile devices (eg, smartphones, smartwatches, and other wearable devices) targeting CVD, diabetes, or cancer. The identified studies were synthesized according to the target disease, the number of enrolled participants and their age, and the study period as well as the DL algorithm used, the main DL outcome, the data set used, the features selected, and the achieved performance.

Results: In total, 20 studies were included in the review. A total of 35% (7/20) of DL studies targeted CVD, 45% (9/20) of studies targeted diabetes, and 20% (4/20) of studies targeted cancer. The most common DL outcome was the diagnosis of the patient's condition for the CVD studies, prediction of blood glucose levels for the studies in diabetes, and early detection of cancer. Most of the DL algorithms used were convolutional neural networks in studies on CVD and cancer and recurrent neural networks in studies on diabetes. The performance of DL was found overall to be satisfactory, reaching >84% accuracy in most studies. In comparison with classic machine learning approaches, DL was found to achieve better performance in almost all studies that reported such comparison outcomes. Most of the studies did not provide details on the explainability of DL outcomes.

Conclusions: The use of DL can facilitate the diagnosis, management, and treatment of major chronic diseases by harnessing mHealth data. Prospective studies are now required to demonstrate the value of applied DL in real-life mHealth tools and interventions.

(JMIR Mhealth Uhealth 2022;10(4):e32344) doi:10.2196/32344

KEYWORDS

mHealth; deep learning; chronic disease; review; mobile phone

Introduction

Background

Chronic, noncommunicable diseases are the leading cause of mortality and disability worldwide. According to the World Health Organization, cardiovascular disease (CVD) is the number 1 cause of death worldwide, taking an estimated 17.9 million lives each year [1]. In 2020, there were approximately 10 million deaths because of cancer [2]. Diabetes is another major chronic disease, with the number of people diagnosed with it increasing dramatically from 108 million in 1980 to 422 million in 2014 [3]. As a consequence of the prevalence of chronic diseases, health care systems around the globe struggle to provide efficient medical care to those patients.

Mobile health (mHealth) has recently emerged as a new paradigm for providing efficient medical care anytime, anywhere. The wide uptake of mobile phones or other mobile electronic communication devices by people has fueled the advancement of their capabilities. Nowadays, mobile devices such as smartphones, smartwatches, and wearable devices can enable robust sensing and processing of health parameters along with communication of health information to patients and caregivers. As a result, they reinforce better daily self-management of chronic diseases by the patients themselves [4] and facilitate remote medical management [5]. In this light, the value of mHealth for chronic diseases has been depicted in several research works [6].

The regular use of mHealth devices around the clock has allowed for the generation of large data sets that can be harnessed by data analytics frameworks toward developing more intelligent mHealth interventions able to identify a range of medical risk factors, improve clinical decision-making, and revolutionize the delivery of health care services [7,8]. The challenge is that the sets of data captured by mHealth devices (eg, sensed data) are often too complex, unstructured, and heterogeneous, thereby creating obstacles in their processing and interpretation through traditional data mining and statistical learning approaches. Deep learning (DL), which is founded on artificial neural networks, appears as a key technology for providing suitable algorithmic frameworks in this direction [9]. DL allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction and requires little engineering by hand [10]. DL models have demonstrated great potential in different domains of health care and have shown excellent performance in computer vision, natural language processing, and mining of electronic health records as well as mHealth modalities and sensor data analytics [11].

Objectives

Despite the potential of DL for mHealth, there have not been targeted reviews in this field. Other reviews have been broad [8,12], not closely related to mHealth [11,13], or not focused on major chronic diseases with the largest prevalence worldwide [14]. In this context, the aim of this paper is to provide a systematic review of the currently available literature and identify recent studies that have used DL based on mHealth data for the diagnosis, prognosis, management, and treatment of major chronic diseases (ie, CVD, diabetes, and cancer). Our ultimate goal is to advance the understanding of researchers, caregivers, and engineers of the progress made in this rapidly developing field.

Methods

Search Strategy

A search was conducted on the web-based bibliographic databases Scopus and PubMed in March 2021 to identify studies published during the last 10 years that incorporated DL in the context of mHealth for CVD, diabetes, and cancer.

Eligibility Criteria

The inclusion criteria for study selection were as follows: (1) DL algorithm or algorithms should be used and quantitative outcomes in terms of their performance should be presented in the study; (2) the DL algorithm in the study should harness mHealth data acquired through a mobile or wearable device; (3) the study should focus on the diagnosis, prognosis, management, or treatment of one of the major chronic diseases with the largest prevalence worldwide (CVD, diabetes, or cancer); and (4) the paper describing the study must have been published in English. Case reports, letters to editors, preprint papers, qualitative studies, surveys or reviews, simulation studies, and studies describing protocols were excluded from the review.

Study Selection

The following string—*(deep learning) OR (neural networks) AND (mobile health) OR (smartphone) OR (mobile phone) OR (mobile device) OR (mobile app) OR (smartwatch) OR (wearable) OR (sensor) AND (health)*—was used for searching within the title, abstract, and keywords of the manuscripts. The retrieved records from Scopus and PubMed were imported into the Mendeley (Mendeley Ltd) bibliography management software to identify duplicates. Authors AT, HK, DK, AK, LK, and AA independently screened the papers that were obtained as a result of the aforementioned search string to minimize bias in the selection process and reduce possible errors. In case of disagreements, these were resolved through discussion between the authors to reach a consensus. The screening procedure took place in 2 stages. In the first stage, the abstracts of the candidate papers for inclusion were screened by the authors according to the defined inclusion and exclusion criteria. In the second stage, the authors read the full manuscripts of the eligible papers, as identified in the first stage, and selected the final papers for inclusion.

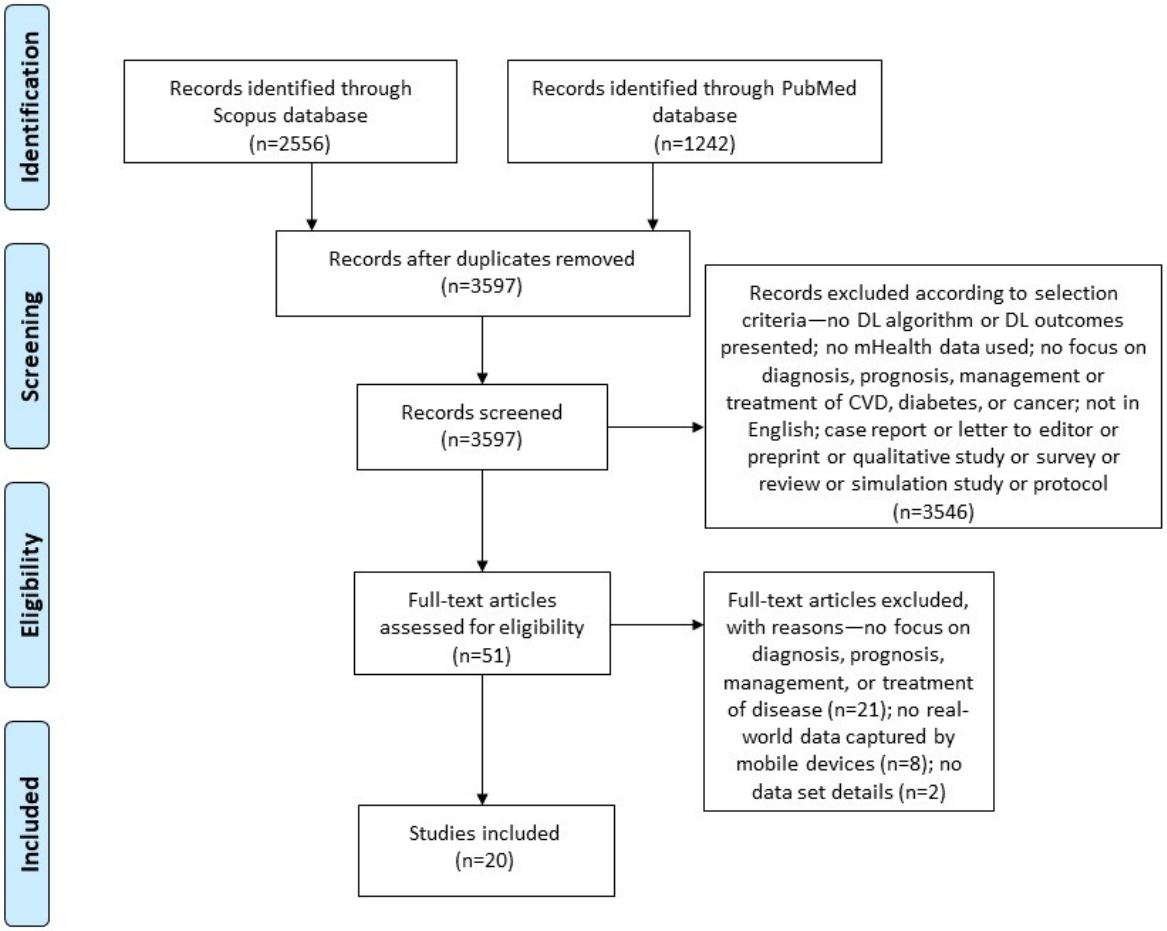
The included studies were synthesized by the authors according to the target disease, the number of enrolled participants and their age, and the study period as well as the DL algorithm used, the main outcome of the algorithm, the data set used, the features selected, and the achieved performance. This systematic review was conducted following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [15]. A completed PRISMA checklist is shown in Multimedia Appendix 1 [16].

Results

Overview

The literature search resulted in 2556 articles from Scopus and 1242 articles from PubMed (3798 articles in total). A total of 94.71% (3597/3798) of records were screened after the removal of 5.29% (201/3798) duplicates. Of those 3597 articles, 3546 (98.58%) were excluded because they did not meet the eligibility criteria. After reading the full texts of the remaining 51 articles, the number of eligible articles was reduced to 20 (39%). Reasons for the exclusion of articles are shown in Figure 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. CVD: cardiovascular disease; DL: deep learning; mHealth: mobile health.



Applications of DL and Outcomes

Overview

Table 1 shows the primary characteristics of the included studies in terms of target disease, number of participants, and their age as well as study duration (where applicable). Of the 20 DL studies, 7 (35%) targeted CVDs, 9 (45%) targeted diabetes, and the remaining 4 (20%) targeted cancer. An interesting finding is that the number of participants included in the DL studies for diabetes was small (range 6-46) compared with CVD (range 10-70,000) and cancer (range 99-917).

Table 1. Characteristics of the included studies (N=20).

Study	Target disease	Participants, N	Age	Study period
Al-Makhadmeh et al [17]	CVD[a]	10	N/A[b]	No
Ali et al [18]	CVD	597 (2 data sets combined with 303 and 294 participants)	29-79 years	No
Dami et al [19]	CVD	Four databases: (1) 70,000 participants, (2) 20,000 participants, (3) 139 patients with hypertension, and (4) 303 participants	N/A	Participants in database 3 were followed for 12 months
Deperlioglu et al [20]	CVD	N/A	N/A	Usability study for 4 months
Fu et al [21]	CVD	20,000	N/A	No (tested in the real world)
Huda et al [22]	CVD	47	N/A	No
Torres-Soto et al [23]	CVD	163	Mean 68 (cardioversion cohort), 56 (exercise stress test cohort), and 67 (ambulatory cohort) years	No
Cappon et al [24]	Diabetes (T1DM[c])	6	20-80 years	8 weeks
Chen et al [25]	Diabetes (T1DM)	6	20-80 years	8 weeks
Efat et al [26]	Diabetes	25	N/A	Data collected during a 2-month period
Faruqui et al [27]	Diabetes (T2DM[d])	10 patients in the smartphone group (overweight or obese)	21-75 years	6 months
Goyal et al [28]	Diabetes	30	N/A	No (tested in the real world)
Joshi et al [29]	Diabetes	46	17-80 years	No
Sánchez-Delacruz et al [30]	Diabetes	15	29-62 years	No
Sevil et al [31]	Diabetes	25	Mean 24.88 (SD 3.15) years	430-hour experiment
Suriyal et al [32]	Diabetes	N/A	N/A	No
Ech-Cherif et al [33]	Cancer	N/A	N/A	No
Guo et al [34]	Cancer	N/A	N/A	No
Hu et al [35]	Cancer	917	N/A	No
Uthoff et al [36]	Cancer	99	Mean 40 (SD 14.1) years	No (tested in the real world)

[a]CVD: cardiovascular disease.

[b]N/A: not applicable.

[c]T1DM: type 1 diabetes mellitus.

[d]T2DM: type 2 diabetes mellitus.

Only 25% (5/20) of the studies reported the integration of their developed model into a DL-empowered system or application that was tested in real life [20,21,28,33,36]. However, none of the studies presented a clinical validation of the deployed systems and applications (eg, through randomized controlled trials).

Approximately 25% (5/20) of the DL studies used an unseen external data set for evaluation purposes to eliminate possible bias and build models that could be generalized [20,23,33,36,37]. Different performance outcomes were reported in the identified studies using DL algorithms (Table 2). None of the included studies used any guidelines for reporting the development and outcomes of the models such as TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) [38]. All the included studies (20/20, 100%) are briefly described below in terms of purpose, data and algorithms used, and evaluation outcomes.

Table 2. Algorithms and outcomes of the included studies (N=20).

Study	DL[a] outcome	DL algorithm	Data used	Features selected	Performance	Comparison with classic ML[b] algorithms
Al-Makhadmeh et al [17]	Detection of heart disease	Higher-order Boltzmann deep belief neural network	123 instances and 23 attributes collected from 10 patients using sensor devices from data sets available in UCI[c] repository	ECG[d], blood pressure, chest pain typology, cholesterol level, vessel information, minimum and maximum heart rate, angina, and depression symptoms	99.5% sensitivity	No
Ali et al [18]	Detection of heart disease	Feedforward network that uses backpropagation techniques and gradient algorithms (ensemble approach)	Cleveland and Hungarian data sets available from UCI repository containing EMR[e] and sensor data (physiological measurements)	Demographic (age and sex), clinical (chest pain type, number of major vessels colored by fluoroscopy, and exercise test results), and sensor (resting blood pressure and fasting blood sugar)	84% accuracy	SVM[f] (71.8%), logistic regression (73.7%), random forest (73.7%), decision tree (74.8%), and naïve Bayes (80.4%)
Dami et al [19]	Prediction of cardiovascular events to prevent SCD[g] and heart attacks	Combination of deep belief network and LSTM[h] RNN[i]	Four databases: (1) Kaggle heart disease data set archive, (2) database from Shahid Beheshti Hospital Research Center, (3) database from PhysioNet site including patients from the Naples Federico II University Hospital in Italy, and (4) UCI4 data set Archive from 1988	Age, sex, weight, height, body surface area, BMI, smoker or not, systolic blood pressure, diastolic blood pressure, intima media thickness, left ventricular mass index, and ejection fraction	88% accuracy, 87% F-measure, and 87% precision	Logistic regression, SVM, and random forest (56% accuracy on average)
Deperlioglu et al [20]	Classification of heart sounds	Autoencoder neural networks	PASCAL[j] B-training heart sound data sets and A-training heart sound data sets	No	96.03% accuracy for normal diagnosis, 91.91% accuracy for extrasystole diagnosis, and 90.11% accuracy for murmur diagnosis	SVM, naïve Bayes, decision tree, and AdaBoost (84.2%-93.3% accuracy)
Fu et al [21]	CVD[k] detection	A hybrid of a CNN[l] and an RNN	The test set includes 15,437 anonymous ECG recordings collected from several tertiary hospitals in China	No	95.53%-99.97% accuracy of CVD	No
Huda et al [22]	Arrhythmia detection	CNN	MIT-BIH[m] arrhythmia data set obtained from PhysioNet	No	94.03% accuracy in classifying abnormal cardiac rhythm	No
Torres-Soto et al [23]	Arrhythmia event detection	Pretraining using convolutional denoising autoencoders followed by CNN, transfer learning, and auxiliary signal quality estimation	Data available through synapse (Synapse ID: syn21985690)	Regions of the upslope from the systolic phase to be informative for AF[n] class-specific predictions	98% sensitivity, 99% specificity, and 96% F_1	Random forest (32% sensitivity, 79% specificity, and 39% F_1)

Study	DL[a] outcome	DL algorithm	Data used	Features selected	Performance	Comparison with classic ML[b] algorithms
Cappon et al [24]	Prediction of short-time blood glucose levels	LSTM RNN	OhioT1DM data set containing CGM[o] data, lifestyle data (diet, exercise, and sleep), galvanic skin response, skin temperature, and magnitude of acceleration	CGM, injected insulin as reported by the pump, and self-reported meals and exercise	RMSE[p] 20.20 and 34.19 for 30- and 60-minute prediction, respectively	No
Chen et al [25]	Prediction of short-time blood glucose levels	Dilated RNNs	The OhioT1DM data set of continuous glucose monitoring data and the corresponding daily events from 6 patients with type 1 diabetes	CGM, insulin doses, carbohydrate intake, and time index; additional data included exercise, heart rate, and skin temperature	15.299 to 22.710 RMSE for different participants	No
Efat et al [26]	Risk level classification of patients with diabetes	Artificial neural network	2-month data from 25 patients with diabetes	Patients' age, sex, sugar level, heart pulse, food intake, sleep time, and exercise or calorie burn	84.29% accuracy, 82.35% sensitivity, and 86.11% specificity	No
Faruqui et al [27]	Prediction of daily glucose levels	LSTM RNN	10 patients with diabetes (T2DM[q]) being overweight or obese	Daily mobile health lifestyle data on diet, physical activity, weight, and previous glucose levels from the day before	33.33% (patient 7) to 86.67% (patient 2) accuracy	KNN[r] regression (10%-56% accuracy)
Goyal et al [28]	Real-time DFU[s] localization	Faster R-CNN[t]	Transfer learning with ImageNet (Stanford Vision Lab) and Microsoft COCO data set; 1775 images of DFUs	Low-level features such as edge detection, corner detection, texture descriptors, shape-based descriptors, and color descriptors	91.8% mean average precision	SVM (70.3% precision)
Joshi et al [29]	Continuous blood glucose monitoring	LMBP[u]	NIR[v] optical spectroscopy data	No	AvgE[w] 6.09%, mARD[x] 6.07%	Multiple polynomial regression—AvgE and mARD were 4.88% and 4.86% for serum glucose examination
Sánchez-Delacruz et al [30]	Diabetic neuropathy detection	Classifiers combined with multilayer perceptron	Raw data from 5 accelerometers	Accelerometer data	85% accuracy	No
Sevil et al [31]	Classification of activity into 5 stages for determining the energy expenditure for diabetes therapy	RNN	Data sets not available	23 selected information features are reported in the paper out of 2216	94.8% classification accuracy	KNN, SVM, naïve Bayes, decision tree, linear discrimination, and ensemble learning (75.7% to 93.1% accuracy)
Suriyal et al [32]	Diabetic retinopathy detection	MobileNets in TensorFlow with the help of RMSprop and asynchronous gradient descent	Data set available in Kaggle database	No	73% accuracy, 74% sensitivity, and 63% specificity	No

Study	DL[a] outcome	DL algorithm	Data used	Features selected	Performance	Comparison with classic ML[b] algorithms
Ech-Cherif et al [33]	Benign and malignant cancer detection	Resource-constrained, mobile-ready deep neural network	Three databases: DermNet, ISIC[y] Archive, and Dermofit Image Library	Cancerous or not	91.33% accuracy	No
Guo et al [34]	Identification of cervix and noncervix images	Ensemble method that consists of 3 DL architectures: RetinaNet, deep SVDD[z], and a customized CNN	Four data sets were used in this study: MobileODT, Kaggle, and COCO2017 for training and validation, and SEVIA[aa] for testing	Normal samples as cervix images and from the anomalous samples as noncervix images	91.6% accuracy and 89% F_1 score	No
Hu et al [35]	Detection of cervical precancer	Automated visual evaluation, RetinaNet, and Adam optimization algorithm	Microsoft COCO images, 7334 training images, 970 validation images, and 1058 test images	No specific features were reported in the paper	ROC[ab] curve (AUC[ac]) of 0.95	No
Uthoff et al [36]	Early detection of precancerous and cancerous lesions in the oral cavity	CNN, VGG-M[ad] network pretrained on the ImageNet data set	170 image pairs	WLI[ae] and AFI[af] provided the most information about type of lesion and size of the affected area	Sensitivity, specificity, positive predictive values, and negative predictive values (81.25%-94.94%); 0.908 AUC	No

[a]DL: deep learning.
[b]ML: machine learning.
[c]UCI: University of California, Irvine.
[d]ECG: electrocardiogram.
[e]EMR: electronic medical record.
[f]SVM: support vector machine.
[g]SCD: sudden cardiac death.
[h]LSTM: long short-term memory.
[i]RNN: recurrent neural network.
[j]PASCAL: Pattern Analysis, Statistical Modeling, and Computational Learning.
[k]CVD: cardiovascular disease.
[l]CNN: convolutional neural network.
[m]MIT-BIH: Massachusetts Institute of Technology–Beth Israel Hospital.
[n]AF: atrial fibrillation.
[o]CGM: continuous glucose monitoring.
[p]RMSE: root mean squared error.
[q]T2DM: type 2 diabetes mellitus.
[r]KNN: k-nearest neighbor.
[s]DFU: diabetic foot ulcer.
[t]R-CNN: region-based convolutional neural network.
[u]LMBP: Levenberg–Marquardt Backpropagation.
[v]NIR: near-infrared.
[w]AvgE: average error.
[x]mARD: mean absolute relative difference.
[y]ISIC: International Skin Imaging Collaboration.
[z]SVDD: support vector data description.
[aa]SEVIA: smartphone-enhanced visual inspection with acetic acid.
[ab]ROC: receiver operating characteristic.
[ac]AUC: area under the curve.
[ad]VGG-M: visual geometry group multi-scale.
[ae]WLI: white-light imaging.
[af]AFI: autofluorescence imaging.

CVD Studies

Huda et al [22] introduced a low-cost, low-power, and wireless electrocardiogram (ECG) monitoring system with DL-based automatic arrhythmia detection. The model was based on a 1D convolutional neural network (CNN) that provided an accuracy of 94.03% in classifying abnormal cardiac rhythm on the Massachusetts Institute of Technology–Beth Israel Hospital Arrhythmia Database.

Deperlioglu et al [20] described a secure Internet of Health Things system to provide real-time support to physicians for the diagnosis of CVDs. Heart sounds were classified using autoencoder neural networks (AENs), and the developed solution demonstrated better results than those reported in the literature studied.

In the study by Ali et al [18], a DL-based ensemble model was used for the detection of heart disease in 597 patients. More specifically, a feedforward neural network was used to perform binary classification of the presence or absence of disease. An 84% accuracy in this classification task was achieved by using publicly available data sets containing sensed data in terms of physiological measurements (such as blood pressure and fasting blood sugar) as well as electronic health record data (including exercise test results, chest pain information, and demographic information).

In the study by Al-Makhadmeh et al [17], the authors proposed the use of a Boltzmann deep belief model to detect whether a patient has heart disease. The model was based on data acquired from 10 patients (publicly available data set), including ECG and blood pressure measurements as well as other diagnostic information such as chest pain and appearance of angina or depression symptoms. A sensitivity of 99.5% was achieved.

Another approach to shed light on the occurrence of arterial and cardiovascular events was examined in the study by Dami et al [19]. A long short-term memory (LSTM) neural network and a deep belief network were used to predict arterial events over the course of a few weeks before the event using ECG recordings and time-frequency features of ECG signals. The proposed LSTM and deep belief network approach had significantly better performance when compared with all other DL approaches and traditional classifications.

Furthermore, in the study by Torres-Soto et al [23], the authors developed DeepBeat, a multitask DL method to detect arrhythmia events for atrial fibrillation in real time using wrist-based photoplethysmography devices. The proposed approach exploited transfer learning, and the resulting models had a sensitivity of 0.98, specificity of 0.99, and F_1 score of 0.93.

In the study by Fu et al [21], an Internet of Things and cloud service system was designed that collected high-quality ECG data and diagnosed 20 types of CVDs using a DL model that was a hybrid between a CNN and a recurrent neural network. The model achieved >0.98 area under the receiver operating characteristic curve score on 17 of the diagnostic items.

Diabetes

For diabetes, there were also several approaches showing impressive performance using DL. For example, in the study by Sevil et al [31], the authors proposed DL with LSTM to determine physical activity states for use in automated insulin delivery systems. The approach exploited a multi-sensor wristband and achieved 94.8% classification accuracy.

In another approach by Suriyal et al [32], DL was used for the detection of diabetic retinopathy using mobile devices for real-time screening without requiring an internet connection. The approach exploited a TensorFlow deep neural network, with a reported accuracy of 73%.

Goyal et al [28] proposed an automated method for the detection and localization of diabetic foot ulcers (DFUs) based on images. The model was robust enough, with a mean average precision of 91.8%, and the trained model could run on simple hardware with a speed of 48 milliseconds for inferencing a single image and with a model size of 57.2 MB. The model was based on transfer learning that was initially trained with ImageNet (Stanford Vision Lab) and Microsoft COCO data sets and with DFU images in the final step. The authors also deployed these models on an Android phone to create real-time object localization for DFUs.

Joshi et al [29] proposed a wearable consumer device called iGLU 2.0, which was based on a DL model for glucose level prediction as a noninvasive, precise, and cost-effective solution to monitor blood glucose levels and control diabetes. The proposed glucometer used the concept of short-wave, near-infrared spectroscopy to predict blood glucose levels. The results were comparable with those of the serum glucose examination, an invasive laboratory examination.

A glucose prediction model was also developed in the study by Chen et al [25]. The authors used a new DL technique based on a dilated recurrent neural network model to predict future glucose levels for a prediction horizon of 30 minutes. Using this model, it was shown that the accuracy of short-time glucose predictions could be significantly improved.

The study by Efat et al [26] introduced a smart health monitoring tool for patients with diabetes. The objective of the authors was to use continuous sensor monitoring and processing with neural networks to provide a continuous evaluation of the patient health risk status.

In the study by Cappon et al [24], an LSTM model for the prediction of blood glucose concentration in patients with type 1 diabetes was proposed. The applied model was based on continuous glucose monitoring data collected from 6 patients as well as insulin dose and self-reported meals and exercise. A root mean squared error of 20.20 for prediction of glucose over the next 30 minutes and of 34.19 for prediction over the next hour was highlighted as the performance outcome of their work.

In the study by Faruqui et al [27], the authors used a DL model based on LSTM and developed a transfer learning strategy (to cope with data scarcity and improve the model's personalization capabilities) to dynamically forecast daily glucose levels. The patient data used for their model were the daily mHealth lifestyle

data and the glucose levels from the day before. The model achieved considerable accuracy in predicting the next day glucose level based on the Clark Error Grid and –10% to +10% range of the actual values on data collected from 10 patients who had been monitored daily for over 6 months.

In the study by Sánchez-Delacruz et al [30], the detection of diabetic neuropathy through the application of a multilayer perceptron combined with additional classifiers on raw accelerometer data was proposed. A total of 15 individuals (10 with diabetic neuropathy and 5 healthy) wearing 5 accelerometers were instructed to walk. The algorithm was able to reach 85% accuracy in diabetic neuropathy recognition.

Cancer

Several studies also focused on cancer. In the study by Hu et al [35], the authors exploited a new DL algorithm called automated visual evaluation for analyzing cervigram images captured by commodity mobile phones to detect cervical precancer. This approach achieved a receiver operating characteristic curve (area under the curve) of 0.95.

In another approach by Uthoff et al [36], the authors used a CNN to enable early detection of precancerous and cancerous lesions in the oral cavity with the potential to reduce morbidity, mortality, and health care costs. To achieve this, the authors used a custom Android app that synchronized an external light-emitting diode and image capture for autofluorescence imaging and white-light imaging on a smartphone. The sensitivity, specificity, positive predictive value, and negative predictive value of the approach ranged from 81.25% to 94.94%.

DL techniques have also been applied for triaging skin cancer detection. The authors in the study by Ech-Cherif et al [33] manually trained a resource-constrained deep CNN called MobileNetV2 to identify the binary classification of skin lesions using benign and malignant as the 2 classes. When the model was tested on an unseen library of images using an iOS mobile app, it was found that all images were correctly classified.

In the study by Guo et al [34], the authors combined the assessment of 3 DL architectures to determine whether an image contained a cervix. The study showed that the ensemble method outperformed individual DL methods. Such data quality algorithms could be used to clean large data sets and provide quality assurance for machine learning (ML) algorithms in routine clinical use.

Architectures of DL Models

DL approaches in mHealth can be efficient by taking advantage of the large volumes of data generated through the use of mobile and sensing devices. In Table 3, we provide details regarding the DL architectures and parameters or hyperparameters used in the selected studies to shed light on the most promising ones used in practice. It is apparent that there is no single best DL architecture to be used for mHealth considering that the selection of the most appropriate DL architecture is mainly data driven.

Regarding the hyperparameters of the studied models, the layers varied between 3 and 50. In most cases, softmax or sigmoid activation functions were used and applied primarily to classification problems, the losses L1 and L2 were <0.01 and, in some cases, Adam optimization was used.

Huda et al [22] proposed a CNN model architecture that consisted of 1D convolution, max-pooling, batch normalization, and dropout layers. The flattened layer output was passed through a fully connected layer with dropout and a second fully connected dense layer. A softmax layer with 14 outputs was then used for arrhythmia classification.

Torres-Soto et al [23] focused on detecting arrhythmia events using unsupervised transfer learning through convolutional denoising autoencoders (CDAEs). The authors applied a 2-stage training to address the unbalanced data problem common to biomedical applications, exploiting a multitask CNN architecture, transfer learning, and an auxiliary signal quality estimation task for atrial fibrillation event detection from spatially segmented physiological photoplethysmography signals. Unsupervised pretraining was performed using CDAEs. The authors then used convolutional and pooling layers in the encoder and upsampling and convolutional layers in the decoder. To obtain the optimal weights, they were randomly initiated according to the He distribution, and the gradient was calculated using the chain rule to backpropagate error derivatives through the decoder network and then through the encoder network. Using a number of hidden units lower than the inputs forces the autoencoder to learn a compressed approximation. The loss function used in pretraining was the mean squared error and was optimized using a backpropagation algorithm. Finally, 3 convolutional layers and 3 pooling layers were used for the encoder segment, and 3 convolutional layers and 3 upsampling layers were used for the decoder segment of the CDAE. A Rectified Linear Unit (ReLU) was applied as the activation function, and Adam was used as the optimization method. Each model was trained with mean squared error loss for 200 epochs, with a reduction in learning rate of 0.001 for every 25 epochs if the validation loss did not improve.

For cervical precancer detection using a smartphone [35], a Resnet-50 architecture was proposed. The whole process started with image augmentation methods (random image scale, random horizontal or vertical flip, random rotation, random shearing, random translation, and transforming the red channel of the image through a γ transformation with γ randomly chosen). Nonmaximum suppression after processing was then followed after cervix or precancerous cervix object detection. The model parameters were initialized with weights pretrained on Microsoft COCO images. All model parameters were then fine-tuned using the visual inspection with acetic acid training data. For the optimization strategy, the authors used the Adam optimization algorithm, fixing the clipnorm parameter at the default of 0.001, and they also used a learning rate of 1×10^{-5}. The metrics used for hyperparameter (number of iterations and batch size) optimization were the mean average precision and validation classification loss.

For smartphone-based oral cancer screening [36], classification using a CNN was applied. For the CNN training, methods commonly used in network training were used, including transfer learning and data augmentation. For data augmentation, the original images were rotated and flipped to feed the network

with more training data. In addition, transfer learning was applied using a visual geometry group multi-scale network pretrained on the ImageNet data set. The network was modified for the task by replacing the final dense layer and softmax layer and then training the network with the available data set.

Goyal et al [28] used transfer learning from massive data sets in nonmedical backgrounds such as ImageNet and Microsoft COCO data sets for the initial training of their image model for DFU localization. The authors used two CNNs, MobileNet and Inception-V2, and set the weight for L2 regularizer as 0.00004 and batch normalization with a decay of 0.9997 and epsilon of 0.001. A batch size of 24 was used along with the optimizer as RMSprop with a learning rate of 0.004 and decay factor of 0.95. The momentum optimizer value was set at 0.9 with a decay of 0.9 and epsilon of 0.1.

The DL approach was used for physical activity classification for automated insulin delivery systems [31], combining different layers including fully connected, LSTM, softmax, regression, ReLU, and dropout layers. In addition, the authors used the L2 regularization term to reduce the risk of overfitting (value 0.05).

In another approach, a TensorFlow deep neural network was used for the detection of diabetic retinopathy [32]. The neural network had 28 convolutional layers and, after each layer, there was a batch normalization and ReLU nonlinear function except for the final layer. The MobileNets training was performed in TensorFlow with the help of RMSprop and asynchronous gradient descent.

Table 3. Model architectures in the included studies (N=20).

Study	DL[a] parameters	DL hyperparameters
Al-Makhadmeh et al [17]	Deep belief network, trained the features using the Boltzmann machine classifiers by computing the energy consumption of the network	Cross-entropy loss of 0.0178, L1 loss of 0.0187, and L2 loss of 0.025
Ali et al [18]	Ensemble DL model composed of 5 layers: the input layer, 3 hidden layers, and the output layer; fully connected hidden layer with 20 nodes	Ada optimizer used and a learning rate with a value of 0.03; ReLU[b] activation function
Dami et al [19]	A deep belief network selected and represented a set of features from the hybrid feature vector and then passed it to the LSTM[c] neural network. The LSTM neural network consists of 5 layers, including input layers, a hidden layer (with 100 hidden units), 2 fully connected layers, a softmax layer, and an output layer	SGD[d] for optimizing cross-entropy as the default loss function
Deperlioglu et al [20]	Autoencoder neural network with a hidden layer size of 10. Softmax layer was used	Scaled conjugate gradient algorithm and cross-entropy cost function was used in the coding layer. The coefficient for the L2 weight *regularizer* was 0.001, the coefficient for the sparsity regularization term was 4, and the sparsity proportion was 0.05
Fu et al [21]	A hybrid of CNN[e] and RNN[f]; 32 convolutional layers (input for CNN) grouped into 8 stages, where each stage was a cascade of four 1D convolutional layers with a kernel size of 16. The final prediction layer was a fully connected dense layer	Before each convolutional layer, a nonlinear transformation occurs, which is a combination of batch normalization, ReLU activation, and a dropout
Huda et al [22]	1D convolution (CNN), max-pooling, and batch normalization. The flattened layer output was passed through a fully connected layer and a second fully connected dense layer. In addition, a softmax layer with 14 outputs was used	Used dropout layers
Torres-Soto et al [23]	Convolutional and pooling layers in the encoder and upsampling and convolutional layers in the decoder; 3 convolutional layers and 3 pooling layers for the encoder segment, and 3 convolutional layers and 3 upsampling layers for the decoder segment of the CDAE[g]	Weights were randomly initiated according to He distribution, and Adam was used as the optimization method. Each model was trained with MSE[h] loss for 200 epochs, with a reduction in learning rate of 0.001 for every 25 epochs if the validation loss did not improve
Cappon et al [24]	A bidirectional LSTM input layer composed of 128 cells having a look-back period of 15 minutes (ie, 3 samples); 2 LSTM layers composed of 64 and 32 cells, respectively; and a fully connected layer consisting of a single neuron computing the BG[i] level prediction at 2 different PHs[j] (ie, 30 and 60 minutes)	BLSTM[k] architecture, hyperparameters, and look-back period were chosen by trial and error to compromise between model complexity and accuracy
Chen et al [25]	A 3-layered DRNN[l] with 32 cells in each layer	1, 2, and 4 dilations implemented for the 3 layers from bottom to top, respectively
Efat et al [26]	RNN	In forward propagation, the sigmoid activation function was applied and, for backpropagation, the margin of error of the output was measured, and the weights were adjusted accordingly
Faruqui et al [27]	LSTM with 5-60 layers and 5-40 number of neurons in the feedforward neural network	Dropout rate of 0.10-0.45. An allowable unit change of 0.01 for the dropout rate parameter and of 1 for the number of neurons in LSTM and feedforward layers was selected. A total of 35 × 55 × 35 = 67,375 combinations were tested before finding the optimal hyperparameters
Goyal et al [28]	Faster R-CNN[m] with ResNet101, Faster R-CNN with Inception-ResnetV2, Faster R-CNN with InceptionV2, and R-FCN[n] with ResNet101	For Faster R-CNN, the weight was set for L2 regularizer as 0.0, initializer that generated a truncated normal distribution with SD of 0.01 and batch normalization with decay of 0.9997 and epsilon of 0.001. For training, a batch size of 2 was used, optimizer as momentum with manual step learning rate and an initial rate of 0.0002, 0.00002 at epoch 40, and 0.000002 at epoch 60. The momentum optimizer value was set at 0.9. For training R-FCN, the same hyperparameters were used as with Faster R-CNN with the only change being in the learning rate set as 0.0005

Study	DL[a] parameters	DL hyperparameters
Joshi et al [29]	DNN[o] with 10 hidden layers	Sigmoid activation functions
Sánchez-Delacruz et al [30]	23 assembled algorithms were tested by combining them with the deep RNA multilayer perceptron. The best results were obtained with the combination of FilteredClassifier and the DL model	The base function was applied to the input values, and the softmax was used as the activation function
Sevil et al [31]	Combination of different layers, including fully connected, LSTM, softmax, regression, ReLU, and dropout layers	L2 regularization=0.05
Suriyal et al [32]	MobileNet CNN with 28 layers. The first layer was a fully connected layer	After each layer, there was batch normalization and a ReLU nonlinear function except at the final layer. Training was done in TensorFlow with the help of RMSprop and asynchronous gradient descent
Ech-Cherif et al [33]	Used the MobileNetV2 model, excluded the classification layer, and replaced it with a dense layer that has two classes: benign and malignant	Used pretrained model MobileNetV2. Adam optimizer was used with a starting learning rate of 0.4. For each experiment, the learning rate was decayed by half every 2 epochs. All experiments were run for 55 epochs. Selected batch size was 32
Guo et al [34]	4 sequentially connected convolutional blocks followed by 2 fully connected layers and softmax for the last layer	N/A[p]
Hu et al [35]	ResNet-50 architecture	Number of iterations and batch size optimization were mean average precision and validation classification loss
Uthoff et al [36]	4 sequentially connected convolutional blocks followed by 2 fully connected layers	N/A

[a]DL: deep learning.
[b]ReLU: Rectified Linear Unit.
[c]LSTM: long short-term memory.
[d]SGD: stochastic gradient descent.
[e]CNN: convolutional neural network.
[f]RNN: recurrent neural network.
[g]CDAE: convolutional denoising autoencoder.
[h]MSE: mean squared error.
[i]BG: blood glucose.
[j]PH: prediction horizon.
[k]BLSTM: bidirectional long short-term memory.
[l]DRNN: dilated recurrent neural network.
[m]R-CNN: region-based convolutional neural network.
[n]R-FCN: region-based fully convolutional network.
[o]DNN: deep neural network.
[p]N/A: not applicable.

Comparison With Classic ML Algorithms

Herein, a presentation of how the DL algorithms used compare with classic ML algorithms, as reported in some of the included studies, is provided. This comparison primarily aims to show whether DL models could bring significant performance gains, which could be critical for their wide adoption by health care providers in routine clinical practice.

In the work by Ali et al [18], the feedforward network for the detection of heart disease based on medical record data and physiological measurements was compared with support vector machine (SVM), random forest, decision tree, and naïve Bayes. The feedforward network achieved 84% accuracy, which was substantially better than the accuracy of classic ML algorithms (72%-80%).

In the work by Dami et al [19], the combination of a deep belief network with LSTM was able to reach 88% accuracy in the prediction of cardiovascular events on data from 4 databases, whereas classic ML algorithms such as logistic regression, SVM, and random forest achieved 56% accuracy on average.

In the paper by Deperlioglu et al [20], AEN was compared thoroughly with additional ML algorithms in other studies. For the PASCAL data set, AEN performed better than all other ML algorithms it was compared with, such as artificial neural networks (82.80-86.50 accuracy), CNN (97.9 accuracy), SVM (90.50 accuracy), naïve Bayes (93.33 accuracy), decision tree (72.76 accuracy), and others. For the PhysioNet data set, AEN performed better than all other ML algorithms, such as CNN (79.50-97.21 accuracy), SVM (83.00 accuracy), wavelet entropy (77.00 accuracy), deep-gated RNA (55.00 accuracy), and others.

DeepBeat in the work by Torres-Soto et al [23] was compared with random forest. However, the sensitivity using random forest was 0.32, the specificity was 0.79, and the F_1 score was 0.39 versus 0.98 sensitivity, 0.99 specificity, and 0.96 F_1 score for the proposed DL methodology.

In the work by Faruqui et al [27], the forecast of daily blood glucose levels through LSTM was achieved with a maximum accuracy of >86% in comparison with the 56% accuracy of k-nearest neighbor regression.

In the work by Goyal et al [28], the localization of DFUs based on imaging data through Faster region-based CNN had a 91.8% average precision. The application of SVM was able to achieve only a 70.3% precision.

Joshi et al [29] applied Levenberg–Marquardt Backpropagation for blood glucose monitoring and achieved an average error of 6.09% in the detection of serum glucose values through near-infrared spectroscopy. However, the use of multiple polynomial regression resulted in a significantly lower average error of 4.88%. This was the only study that showed that a classic ML approach was better than a DL approach.

In the work by Sevil et al [31], the authors compared the performance of their recurrent neural network with k-nearest neighbor, regression SVMs, decision trees, naïve Bayes, Gaussian process regression, ensemble learning, and linear discrimination and regression, which achieved 75.7% to 93.1% accuracy, whereas the proposed approach achieved a classification accuracy of 94.8%.

Explainability Aspects

When developing models for decision support, there is a need to provide transparent and trustworthy models able to produce not only reliable but also explainable predictions [39]. However, a known problem with DL models is that they lack interpretability and explainability, which hinders their wide adoption in clinical practice.

Explainability deals with the implementation of transparency and traceability of statistical black - box ML methods. Although attempts to tackle problems related to explanation and interpretability have existed for several years now, there has been an exceptional growth in research efforts in the last couple of years [40]. Approaches for explainability include keeping track of how algorithms are used, which features are the most important for predicting the target variable, and how the algorithm used can be improved, thereby providing hints and clues to guide further developments and enabling the detection of erroneous reasoning through techniques of advanced visualization and signal processing. The challenge is hard as explanations should be sound and complete in statistical and causal terms and yet comprehensible to users, subject to decisions.

This difficulty is also demonstrated in the presented works under review in several cases (eg, in the study by Guo et al [34]). The authors visually analyzed the error cases to better understand why the results were wrong. In only 5% (1/20) of the studies [24], the authors exploited Shapley Additive Explanations (ie, a newly developed approach to interpret DL model predictions [41]). Focusing on predicting glucose concentration in type 1 diabetes, the Shapley Additive Explanations identified that high values of continuous glucose measurements resulted in high predicted blood glucose levels and that high insulin negatively affected the model output.

Discussion

Principal Findings

This work presented a systematic literature review of the applications of DL in mHealth for three major chronic diseases that pose a significant international burden: CVD, diabetes, and cancer. To the authors' knowledge, this is the first systematic review of DL in mHealth for these diseases. The principal outcome of this review is that DL approaches have been used effectively for a variety of diagnostic and predictive tasks in mHealth. More specifically, the most common DL outcomes were found to be (1) diagnosis of the patient's condition for CVDs, (2) prediction of blood glucose levels for diabetes, and (3) early detection of cancer.

CNNs and recurrent neural networks were the DL algorithms used in most studies. It is worth mentioning that CNNs have been successfully applied to deal with not only computer vision medical tasks but also other tasks based on nonimaging data, such as detection of arrhythmia [22,23] or CVD [21]. Overall, the performance of DL approaches was found to be satisfactory considering that >84% accuracy was achieved in most studies.

In comparison with classic ML approaches, DL was found to achieve better performance in almost all studies that reported such comparison outcomes. This finding shows the value and potential of DL in mHealth for realizing highly intelligent mHealth systems and interventions that could significantly improve clinical decision-making processes. Nevertheless, the authors of this paper acknowledge that DL models require more effort compared with ML models for the preprocessing part, especially when the architecture is based on transfer learning, a common method in most of the image-processing architectures.

The diversity of the identified DL models in the mHealth studies confirms that, for the selection of the most appropriate DL architecture, the one-size-fits-all approach does not apply, a finding that has also been indicated in DL reviews for other fields [42,43]. Another remark is that the architectures of the models in the mHealth studies, as well as the methodologies used for training, were not stated in a consistent manner. This renders the comparison of various approaches between works nontrivial for the interested researcher. None of the included studies used guidelines for reporting the development or outcomes of the models, which could have facilitated the assessment and interpretation of their findings [44].

Most of the included studies dealt with the retrospective technical validation of DL approaches. More thorough external validation is required to prove the generalizability of the DL findings considering that only a minority of the DL studies used an unseen external data set for evaluation purposes. Furthermore, no randomized controlled trials or other types of clinical validation studies with intelligent digital health interventions relying on DL approaches were found in this review [45]. In

this respect, further work by the research community is needed to develop DL-empowered systems and applications and prove their clinical effectiveness in health care settings within prospective clinical studies.

Although DL was found to be an effective approach in mHealth for chronic diseases, the explainability of DL outcomes has been scarce. It is apparent that future work is required on the explainability of the DL models developed for chronic diseases as only 5% (1/20) of the studies in this review considered this important dimension [24]. Leveraging explainable models would enhance trust in artificial intelligence and help clinicians make informed judgments [46,47], thereby promoting the real-life use of those models in daily clinical practice. Equally important for the developed models is to support their fairness by ensuring that they mitigate inequalities between individuals and groups of individuals, in particular differences in sex or gender, age, ethnicity, income, education, and geography. In the reviewed studies, mitigation of differences was missing in most cases, merely because of the lack of adequate data. However, if DL models are to be used in daily practice, they should also guarantee fairness and universality [48,49].

Limitations

This review should be interpreted within the context of its limitations. The authors used a limited set of terms for the search of the literature, including keywords such as *DL* and *neural networks*, combined with keywords related to mHealth. Keywords for specific DL algorithms were not used. This might have inadvertently omitted studies that could have contributed to the progress made in DL applications for mHealth. Articles were searched in a limited number of databases (ie, PubMed and Scopus); two of the most widely used databases internationally nonetheless. No hand search was conducted on any studies reported in other reviews or the included studies, and there was no assessment of the interrater reliability between the authors. A meta-analysis was not possible because of the heterogeneity of the included studies. On the basis of the selected inclusion and exclusion criteria, a small number of eligible studies were included and examined in this review, which limits the generalizability of the findings.

Conclusions

This review found that DL approaches for chronic diseases could be the vehicle for the translation of big mHealth data into useful knowledge about patient health. Nevertheless, to unlock the full potential of DL, the research community needs to move beyond the conduction of retrospective validation studies and provide robust evidence of the added clinical value of DL-based tools in real-life settings compared with standard care.

Acknowledgments

The work presented in this paper was supported by the Center for eHealth Applications and Services, Institute of Computer Science, Foundation for Research and Technology Hellas. Authors AT, AA, KV, and DT were supported by the Horizon 2020 research and innovation program of the European Union under grant agreements 945246 (DigiCare4You) and 727409 (DM4ALL-PROEMPOWER).

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2009 checklist.
[DOC File , 121 KB - mhealth_v10i4e32344_app1.doc]

References

1. Noncommunicable diseases. World Health Organization. URL: http://www.who.int/mediacentre/factsheets/fs355/en/ [accessed 2022-03-27]
2. Ferlay J, Colombet M, Soerjomataram I, Parkin D, Piñeros M, Znaor A, et al. Cancer statistics for the year 2020: an overview. Int J Cancer 2021 Apr 05:778-789. [doi: 10.1002/ijc.33588] [Medline: 33818764]
3. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. Lancet 2016 Apr 09;387(10027):1513-1530 [FREE Full text] [doi: 10.1016/S0140-6736(16)00618-8] [Medline: 27061677]
4. Triantafyllidis AK, Koutkias VG, Chouvarda I, Maglaveras N. A pervasive health system integrating patient monitoring, status logging, and social sharing. IEEE J Biomed Health Inform 2013 Jan;17(1):30-37. [doi: 10.1109/titb.2012.2227269]
5. Seto E, Leonard KJ, Cafazzo JA, Barnsley J, Masino C, Ross HJ. Perceptions and experiences of heart failure patients and clinicians on the use of mobile phone-based telemonitoring. J Med Internet Res 2012 Feb 10;14(1):e25 [FREE Full text] [doi: 10.2196/jmir.1912] [Medline: 22328237]
6. Triantafyllidis A, Kondylakis H, Votis K, Tzovaras D, Maglaveras N, Rahimi K. Features, outcomes, and challenges in mobile health interventions for patients living with chronic diseases: a review of systematic reviews. Int J Med Inform 2019 Dec;132:103984. [doi: 10.1016/j.ijmedinf.2019.103984] [Medline: 31605884]

7. Baladrón C, Gómez de Diego JJ, Amat-Santos IJ. Big data and new information technology: what cardiologists need to know. Rev Esp Cardiol (Engl Ed) 2021 Jan;74(1):81-89. [doi: 10.1016/j.rec.2020.06.036] [Medline: 33008773]
8. Triantafyllidis AK, Tsanas A. Applications of machine learning in real-life digital health interventions: review of the literature. J Med Internet Res 2019 Apr 05;21(4):e12286 [FREE Full text] [doi: 10.2196/12286] [Medline: 30950797]
9. Lane N, Georgiev P. Can deep learning revolutionize mobile sensing? In: Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications. 2015 Presented at: HotMobile '15: The 16th International Workshop on Mobile Computing Systems and Applications; Feb 12 - 13, 2015; Santa Fe New Mexico USA. [doi: 10.1145/2699343.2699349]
10. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]
11. Ravi D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, et al. Deep learning for health informatics. IEEE J Biomed Health Inform 2017 Jan;21(1):4-21. [doi: 10.1109/jbhi.2016.2636665]
12. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Brief Bioinform 2018 Nov 27;19(6):1236-1246 [FREE Full text] [doi: 10.1093/bib/bbx044] [Medline: 28481991]
13. Xiao C, Choi E, Sun J. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. J Am Med Inform Assoc 2018 Oct 01;25(10):1419-1428 [FREE Full text] [doi: 10.1093/jamia/ocy068] [Medline: 29893864]
14. Valliani A, Ranti D, Oermann E. Deep learning and neurology: a systematic review. Neurol Ther 2019 Dec;8(2):351-365 [FREE Full text] [doi: 10.1007/s40120-019-00153-8] [Medline: 31435868]
15. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009 Jul 21;339(jul21 1):b2700 [FREE Full text] [doi: 10.1136/bmj.b2700] [Medline: 19622552]
16. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
17. Al-Makhadmeh Z, Tolba A. Utilizing IoT wearable medical device for heart disease prediction using higher order Boltzmann model: a classification approach. Measurement 2019 Dec;147:106815. [doi: 10.1016/j.measurement.2019.07.043]
18. Ali F, El-Sappagh S, Islam SR, Kwak D, Ali A, Imran M, et al. A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion. Inf Fusion 2020 Nov;63:208-222. [doi: 10.1016/j.inffus.2020.06.008]
19. Dami S, Yahaghizadeh M. Predicting cardiovascular events with deep learning approach in the context of the internet of things. Neural Comput Applic 2021 Jan 03;33(13):7979-7996. [doi: 10.1007/s00521-020-05542-x]
20. Deperlioglu O, Kose U, Gupta D, Khanna A, Sangaiah AK. Diagnosis of heart diseases by a secure internet of health things system based on autoencoder deep neural network. Comput Commun 2020 Oct 01;162:31-50 [FREE Full text] [doi: 10.1016/j.comcom.2020.08.011] [Medline: 32843778]
21. Fu Z, Hong S, Zhang R, Du S. Artificial-intelligence-enhanced mobile system for cardiovascular health management. Sensors (Basel) 2021 Jan 24;21(3):773 [FREE Full text] [doi: 10.3390/s21030773] [Medline: 33498892]
22. Huda N, Khan S, Abid R, Shuvo S, Labib M, Hasan T. A low-cost, low-energy wearable ECG system with cloud-based arrhythmia detection. In: Proceedings of the 2020 IEEE Region 10 Symposium (TENSYMP). 2020 Presented at: 2020 IEEE Region 10 Symposium (TENSYMP); Jun 5-7, 2020; Dhaka, Bangladesh. [doi: 10.1109/tensymp50017.2020.9230619]
23. Torres-Soto J, Ashley EA. Multi-task deep learning for cardiac rhythm detection in wearable devices. NPJ Digit Med 2020;3:116 [FREE Full text] [doi: 10.1038/s41746-020-00320-4] [Medline: 32964139]
24. Cappon G, Meneghetti L, Prendin F, Pavan J, Sparacino G, Favero S, et al. A personalized and interpretable deep learning based approach to predict blood glucose concentration in type 1 diabetes. In: Proceedings of the 5th International Workshop on Knowledge Discovery in Healthcare Data, KDH 2020. 2020 Presented at: 5th International Workshop on Knowledge Discovery in Healthcare Data, KDH 2020; Aug 29-30, 2020; Spain URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85093844631&partnerID=40&md5=1f91df5cc25e0eb0abb73f72ccfd6eb8
25. Chen J, Li K, Herrero P, Zhu T, Georgiou P. Dilated recurrent neural network for short-time prediction of glucose concentration. In: Proceedings of the 3rd International Workshop on Knowledge Discovery in Healthcare Data. 2018 Presented at: 3rd International Workshop on Knowledge Discovery in Healthcare Data; Jul 13, 2018; Stockholm, Sweden URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85051009758&partnerID=40&md5=61d3bd966e5e3faeb09bda20799a87d9
26. Efat M, Rahman S, Rahman T. IoT based smart health monitoring system for diabetes patients using neural network. In: Cyber Security and Computer Science. Cham: Springer; 2020.
27. Faruqui SH, Du Y, Meka R, Alaeddini A, Li C, Shirinkam S, et al. Development of a deep learning model for dynamic forecasting of blood glucose level for type 2 diabetes mellitus: secondary analysis of a randomized controlled trial. JMIR Mhealth Uhealth 2019 Nov 01;7(11):e14452 [FREE Full text] [doi: 10.2196/14452] [Medline: 31682586]

28. Goyal M, Reeves ND, Rajbhandari S, Yap MH. Robust methods for real-time diabetic foot ulcer detection and localization on mobile devices. IEEE J Biomed Health Inform 2019 Jul;23(4):1730-1741. [doi: 10.1109/JBHI.2018.2868656] [Medline: 30188841]
29. Joshi AM, Jain P, Mohanty SP, Agrawal N. iGLU 2.0: a new wearable for accurate non-invasive continuous serum glucose measurement in IoMT framework. IEEE Trans Consum Electron 2020 Nov;66(4):327-335. [doi: 10.1109/tce.2020.3011966]
30. Sánchez-DelaCruz E, Weber R, Biswal RR, Mejía J, Hernández-Chan G, Gómez-Pozos H. Gait biomarkers classification by combining assembled algorithms and deep learning: results of a local study. Comput Math Methods Med 2019 Dec 19;2019:3515268-3515214 [FREE Full text] [doi: 10.1155/2019/3515268] [Medline: 31933676]
31. Sevil M, Rashid M, Maloney Z, Hajizadeh I, Samadi S, Askari MR, et al. Determining physical activity characteristics from wristband data for use in automated insulin delivery systems. IEEE Sens J 2020 Nov;20(21):12859-12870 [FREE Full text] [doi: 10.1109/jsen.2020.3000772] [Medline: 33100923]
32. Suriyal S, Druzgalski C, Gautam K. Mobile assisted diabetic retinopathy detection using deep neural network. In: Proceedings of the 2018 Global Medical Engineering Physics Exchanges/Pan American Health Care Exchanges (GMEPE/PAHCE). 2018 Presented at: 2018 Global Medical Engineering Physics Exchanges/Pan American Health Care Exchanges (GMEPE/PAHCE); Mar 19-24, 2018; Porto, Portugal. [doi: 10.1109/gmepe-pahce.2018.8400760]
33. Ech-Cherif A, Misbhauddin M, Ech-Cherif M. Deep neural network based mobile dermoscopy application for triaging skin cancer detection. In: Proceedings of the 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS). 2019 Presented at: 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS); May 1-3, 2019; Riyadh, Saudi Arabia. [doi: 10.1109/cais.2019.8769517]
34. Guo P, Xue Z, Mtema Z, Yeates K, Ginsburg O, Demarco M, et al. Ensemble deep learning for cervix image selection toward improving reliability in automated cervical precancer screening. Diagnostics (Basel) 2020 Jul 03;10(7):451 [FREE Full text] [doi: 10.3390/diagnostics10070451] [Medline: 32635269]
35. Hu L, Horning M, Banik D, Ajenifuja O, Adepiti C, Yeates K, et al. Deep learning-based image evaluation for cervical precancer screening with a smartphone targeting low resource settings - Engineering approach. In: Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). 2020 Presented at: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); Jul 20-24,2020; Montreal, QC, Canada. [doi: 10.1109/embc44109.2020.9175863]
36. Uthoff RD, Song B, Sunny S, Patrick S, Suresh A, Kolur T, et al. Point-of-care, smartphone-based, dual-modality, dual-view, oral cancer screening device with neural network classification for low-resource communities. PLoS One 2018 Dec 5;13(12):e0207493 [FREE Full text] [doi: 10.1371/journal.pone.0207493] [Medline: 30517120]
37. Mattioli G, Wong M, Angotti R, Mazzola C, Arrigo S, Gandullia P, et al. Total oesophago-gastric dissociation in neurologically impaired children: laparoscopic vs robotic approach. Int J Med Robot 2020 Feb;16(1):e2048. [doi: 10.1002/rcs.2048] [Medline: 31797517]
38. Collins G, Reitsma J, Altman D, Moons K. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD Statement. Br J Surg 2015 Feb;102(3):148-158. [doi: 10.1002/bjs.9736] [Medline: 25627261]
39. Ahmad M, Teredesai A, Eckert C. Interpretable machine learning in healthcare. In: Proceedings of the 2018 IEEE International Conference on Healthcare Informatics (ICHI). 2018 Presented at: 2018 IEEE International Conference on Healthcare Informatics (ICHI); Jun 4-7, 2018; New York, NY, USA. [doi: 10.1109/ichi.2018.00095]
40. Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D. A survey of methods for explaining black box models. ACM Comput Surv 2019 Sep 30;51(5):1-42. [doi: 10.1145/3236009]
41. A game theoretic approach to explain the output of any machine learning model. GitHub. URL: https://github.com/slundberg/shap [accessed 2022-03-27]
42. Khamparia A, Singh KM. A systematic review on deep learning architectures and applications. Expert Syst 2019 Mar 27;36(3):e12400. [doi: 10.1111/exsy.12400]
43. Hernández-Blanco A, Herrera-Flores B, Tomás D, Navarro-Colorado B. A systematic review of deep learning approaches to educational data mining. Complexity 2019 May 12;2019:1-22. [doi: 10.1155/2019/1306039]
44. Yusuf M, Atal I, Li J, Smith P, Ravaud P, Fergie M, et al. Reporting quality of studies using machine learning models for medical diagnosis: a systematic review. BMJ Open 2020 Mar 23;10(3):e034568 [FREE Full text] [doi: 10.1136/bmjopen-2019-034568] [Medline: 32205374]
45. Buijink AW, Visser BJ, Marshall L. Medical apps for smartphones: lack of evidence undermines quality and safety. Evid Based Med 2013 Jun 25;18(3):90-92. [doi: 10.1136/eb-2012-100885] [Medline: 22923708]
46. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. J Med Internet Res 2020 Jun 19;22(6):e15154 [FREE Full text] [doi: 10.2196/15154] [Medline: 32558657]
47. Rai A. Explainable AI: from black box to glass box. J Acad Mark Sci 2019 Dec 17;48(1):137-141. [doi: 10.1007/s11747-019-00710-5]
48. Stratigi M, Kondylakis H, Stefanidis K. Fairness in group recommendations in the health domain. In: Proceedings of the 2017 IEEE 33rd International Conference on Data Engineering (ICDE). 2017 Presented at: 2017 IEEE 33rd International Conference on Data Engineering (ICDE); Apr 19-22, 2017; San Diego, CA, USA. [doi: 10.1109/icde.2017.217]

49. Stratigi M, Kondylakis H, Stefanidis K. FairGRecs: fair group recommendations by exploiting personal health information. In: Database and Expert Systems Applications. Cham: Springer; 2018.

Abbreviations

AEN: autoencoder neural network
CDAE: convolutional denoising autoencoder
CNN: convolutional neural network
CVD: cardiovascular disease
DFU: diabetic foot ulcer
DL: deep learning
ECG: electrocardiogram
LSTM: long short-term memory
mHealth: mobile health
ML: machine learning
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
ReLU: Rectified Linear Unit
SVM: support vector machine
TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis

Edited by L Buis; submitted 23.07.21; peer-reviewed by G Cappon, S Faruqui, S Hong; comments to author 15.12.21; revised version received 26.01.22; accepted 22.02.22; published 04.04.22.

Please cite as:
Triantafyllidis A, Kondylakis H, Katehakis D, Kouroubali A, Koumakis L, Marias K, Alexiadis A, Votis K, Tzovaras D
Deep Learning in mHealth for Cardiovascular Disease, Diabetes, and Cancer: Systematic Review
JMIR Mhealth Uhealth 2022;10(4):e32344
URL: https://mhealth.jmir.org/2022/4/e32344
doi:10.2196/32344
PMID:35377325

Review

The Implementation of Behavior Change Techniques in mHealth Apps for Sleep: Systematic Review

Amber Carmen Arroyo[1], PhD; Matthew J Zawadzki[1], PhD

Department of Psychological Sciences, University of California, Merced, CA, United States

Corresponding Author:
Amber Carmen Arroyo, PhD
Department of Psychological Sciences
University of California
5200 N Lake Road
Merced, CA, 95343
United States
Phone: 1 209 228 4787
Email: aarroyo22@ucmerced.edu

Abstract

Background: Mobile health (mHealth) apps targeting health behaviors using behavior change techniques (BCTs) have been successful in promoting healthy behaviors; however, their efficacy with sleep is unclear. Some work has shown success in promoting sleep through mHealth, whereas there have been reports that sleep apps can be adverse and lead to unhealthy obsessions with achieving perfect sleep.

Objective: This study aims to report and describe the use of BCTs in mHealth apps for sleep with the following research questions: How many BCTs are used on average in sleep apps, and does this relate to their effectiveness on sleep outcomes? Are there specific BCTs used more or less often in sleep apps, and does this relate to their effectiveness on sleep outcomes? Does the effect of mHealth app interventions on sleep change when distinguishing between dimension and measurement of sleep?

Methods: We conducted a systematic review following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines to review articles on mHealth app interventions for sleep published between 2010 and 2020.

Results: A total of 12 studies met the eligibility criteria. Most studies reported positive sleep outcomes, and there were no negative effects reported. Sleep quality was the most common dimension of sleep targeted. Subjective measures of sleep were used across all apps, whereas objective measures were often assessed but rarely reported as part of results. The average number of BCTs used was 7.67 (SD 2.32; range 3-11) of 16. Of the 12 studies, the most commonly used BCTs were feedback and monitoring (n=11, 92%), shaping knowledge (n=11, 92%), goals and planning (n=10, 83%), and antecedents (n=10, 83%), whereas the least common were scheduled consequences (n=0, 0%), self-belief (n=0, 0%), and covert learning (n=0, 0%). Most apps used a similar set of BCTs that unfortunately did not allow us to distinguish which BCTs were present when studies reported more positive outcomes.

Conclusions: Our study describes the peer-reviewed literature on sleep apps and provides a foundation for further examination and optimization of BCTs used in mHealth apps for sleep. We found strong evidence that mHealth apps are effective in improving sleep, and the potential reasons for the lack of adverse sleep outcome reporting are discussed. We found evidence that the type of BCTs used in mHealth apps for sleep differed from other health outcomes, although more research is needed to understand how BCTs can be implemented effectively to improve sleep using mHealth and the mechanisms of action through which they are effective (eg, self-efficacy, social norms, and attitudes).

(JMIR Mhealth Uhealth 2022;10(4):e33527) doi:10.2196/33527

KEYWORDS

behavior change techniques; sleep; mHealth; apps; digital health; mobile phone

Introduction

Background

Mobile health (mHealth) is the use of mobile technology (eg, smartphones) to improve health practices. mHealth interventions have incredible potential to implement large-scale health interventions at low cost, and their efficacy to promote health behaviors such as physical activity and diet has been established [1-4]. Despite the increasing awareness that sleep has a critical association with the development and progression of the largest killers in the United States (eg, heart disease, cancer, stroke, Alzheimer disease, and diabetes) [5,6] sleep has not been studied to the same extent as other health behaviors [7]. Notably, the use of mHealth apps for sleep across peer-reviewed studies has not been systematically reviewed.

There is a growing concern that apps should not be used for sleep because they can cause orthosomnia: an individual's unhealthy obsession with achieving perfect sleep [8]. However, there are other reports that technology has the potential to improve sleep outcomes [9-11]. Possibly, the lack of consensus for the viability of mHealth apps with sleep is because of the components making up the intervention. The contents of all interventions are known as behavior change techniques (BCTs), which are designed to change or redirect the determinants that regulate behavior [12]. This study aims to examine the use of BCTs across mHealth apps for sleep to identify best practices for future mHealth app intervention development.

BCTs in mHealth Apps for Sleep

Overview

BCTs are the irreducible active ingredients of all interventions used to facilitate behavior change [12]. The most common classification of BCTs is the *BCT Taxonomy V1* [12] and is widely considered the gold standard for behavior change research design and reporting [12-14]. The taxonomy defines 16 BCT clusters, with each representing a principal method of behavior change. Incorporating these evidence-based BCTs in interventions is recommended because they are known to successfully change health behaviors [15].

The efficacy of BCTs in mHealth apps for sleep has not been systematically examined despite some initial promise on their efficacy. One study did evaluate the number and type of BCTs used in commercially available sleep apps, reporting a greater number of BCTs used for physical activity compared with sleep, although sleep had a larger number of BCTs used than sedentary behavior [1]. However, they did not examine whether the number of BCTs was associated with better or worse treatment outcomes. The researchers also reported some of the BCTs used in mHealth apps for sleep were dissimilar to those seen in apps for physical activity (ie, social support and reward and threat), which may reflect a conscious or unconscious understanding by app developers that sleep is a unique construct. One particular feature that differentiates sleep from other health behaviors is that it is not always under an individual's control. For example, an able-bodied individual can usually control how many steps they take in a day or how many calories they eat but cannot directly determine how many hours of sleep they get in a day or how many times they wake up in the middle of the night. As such, there may be some BCTs that are uniquely suited for mHealth apps for sleep. Although there is no overall framework for which BCTs might work best with sleep in mHealth apps, a brief overview is provided in the subsequent section to hypothesize possible associations.

Hypothesizing Differences in BCTs for Sleep

The implementation of BCTs that are aimed at changing aspects of behavior before sleep could be more appropriate for sleep than BCTs focused on future outcomes or consequences of sleep. This is because directing attention to the outcomes or consequences of sleep necessitates an anticipation of future events (ie, worry), which is often accompanied by anxiety. Anxiety is known to interfere with the successful initiation of sleep [16,17], as sleep often requires a quiet state of mind to be achieved and performed successfully [18]. Anxiety is also related to physiological arousal that disrupts the relaxation process needed for sleep [19-21]. Therefore, the BCTs that direct attention to the predictors of sleep may be optimal for sleep apps because they may bypass the arousal-related processes that could interfere with the initiation of sleep. BCTs focused on aspects of behavior before sleep are shaping knowledge, associations, repetition and substitution, antecedents, regulation, and self-belief. For example, antecedents can work to restructure or add objects to the physical environment to promote the behavior, such as creating a sleep sanctuary in one's bedroom (eg, installing blackout curtains, white noise machine, or comfortable bedding). The BCT shaping knowledge can work to increase knowledge or skills to perform the behavior, such as providing information about antecedents that facilitate or harm sleep (eg, avoid caffeine within 6 hours of bed or avoid vigorous exercise within 2 hours of bed).

In opposition, the BCTs that direct an individual's attention to the outcomes or consequences of sleep are natural consequences, comparison of outcomes, reward and threat, scheduled consequences, and covert learning. The use of these BCTs may be less appropriate for sleep. For example, natural consequences emphasizes the consequences of a behavior, such as highlighting the health consequences of inadequate sleep duration to discourage poor sleep practices, but could have the unintended effect of increasing anxiety surrounding sleep duration and inhibit one's ability to relax enough to fall asleep. This may even be the case with positive consequences, such as the reward in the reward and threat BCT. For instance, if an individual is told they will receive a reward if they obtain 7 to 9 hours of sleep for an entire week, it may result in pressure or anxiety surrounding sleep and interfere with sleep onset or maintenance.

There are some BCTs that do not fit clearly into either category and therefore no hypotheses will be made for their frequency of use in mHealth apps for sleep. These BCTs include social support, comparison of behavior, goal setting, feedback and monitoring, and identity. More information on BCTs and examples of their implementation with sleep are provided in Multimedia Appendix 1.

Sleep Outcome Measures

When examining the use of BCTs in mHealth app interventions for sleep, it is important to consider how sleep is operationalized. It is possible that the way sleep outcomes have been measured in the previous research may partially explain why some sleep apps appear to be beneficial, whereas others appear to be harmful [22]. One common and well-supported approach to studying sleep health identifies five dimensions of sleep [23]: sleep quality (satisfaction with sleep), sleep duration (total amount of sleep acquired over a 24-hour period), sleep continuity (ease of falling asleep and staying asleep), sleep timing (placement of sleep in a 24-hour period), and sleepiness (ability to maintain wakefulness). Researchers have been advised to evaluate multiple dimensions of sleep concurrently to obtain a more accurate representation of an individual's sleep health. For example, if an intervention increases sleep duration but sleep continuity decreases (ie, efficiency of time in bed in relation to time asleep), then the intervention may not be considered successful. Indeed, there are cases of interventions improving some dimensions of sleep whereas others remain the same (eg, sleep quality improves but sleep duration remains the same) [24]. The dimensions of sleep also have different associations with health [23], suggesting they are unique constructs and should be treated as such in reviews of the literature. It is therefore important to consider the dimension of sleep being targeted while examining mHealth app interventions for sleep.

In addition to sleep dimensions, the extent to which sleep health is subjectively or objectively measured may also be an important consideration. Subjective sleep is a self-reported appraisal of how one is sleeping and is assessed with retrospective questionnaires and sleep diaries [23]. Objective sleep is a measured observation of sleep parameters that is not directly controlled by the participant and is often assessed with remote behavioral and physiological technology [23]. There is a growing body of research suggesting that subjective and objective sleep measures may not be redundant. For example, multiple studies have reported that their subjective and objective measures of sleep continuity (specifically, the number of awakenings in a night) did not correlate with one another [25-27]. There have also been cases of individuals reporting subjectively defined insomnia but objectively normal sleep [28]. In addition, it has been noted that subjective and objective measures of sleep can differentially predict treatment efficacy in interventions for insomnia [29]. It is therefore recommended to use both subjective and objective methods of sleep measurement to obtain a comprehensive understanding of how an individual is sleeping [25,30]. This study will examine both sleep overall and also will distinguish between sleep dimension, and whether sleep outcomes were subjectively or objectively measured.

This Study

The aim of this study is to report and describe the use of BCTs in mHealth app interventions for sleep in the peer-reviewed literature. Despite several systematic reviews examining mHealth interventions for chronic disease management and other health behaviors [2], there has not been a systematic review dedicated to examining BCTs in fully automated mHealth apps for sleep. One systematic review [9] evaluated the design engineering and implementation of mHealth apps for sleep disturbances but included papers with no quantitative evaluations of sleep (eg, apps that only measured and tracked sleep) and apps that required clinician input (ie, did not function autonomously). There have also been promising reviews on the efficacy of internet-delivered interventions for insomnia [31] but they notably did not review mobile apps that offer unique features including portability, touchscreen interactivity, and notifications and alerts [32].

Given the rapidly expanding public health issue of inadequate sleep [33] and the vast number of commercially available sleep apps compared with the small number of peer-reviewed studies of sleep apps [9], conducting a systematic review of fully automated mHealth apps for sleep in the peer-reviewed literature is required. In this systematic review, we aim to identify when and how BCTs were used in mHealth apps for sleep, with the hope of informing future app development and providing areas to focus on in future meta-analyses. We had three research questions (RQs):

1. How many BCTs are used on average in sleep apps, and does this relate to their effectiveness with sleep?
2. Are there specific BCTs used more or less often in sleep apps, and does this relate to their effectiveness with sleep?
3. Does the effect of mHealth app interventions on sleep change when distinguishing between dimension and measurement of sleep?

Methods

Article Inclusion Procedure

Search Strategy

The search strategy and study selection methods were adopted from prior research reviewing the effectiveness of mobile phone apps in achieving behavior change for a broad range of health behaviors [34]. A search of the PubMed database included all articles published between January 1, 2010, and January 1, 2020. The 2010 start date was specified to acknowledge that the creation of smartphone apps was relatively recent [35,36]. The PubMed database was chosen because of its strong usability for systematic reviews [36] and given the substantial portion of sleep studies occurring in medical research. The search string was specified as *Title [sleep* OR insomnia*] AND Title/Abstract [smartphone OR phone OR mHealth OR eHealth OR telehealth OR mobile OR digital OR iPhone OR Android] AND Title/Abstract [CBT OR cognitive behavioral OR health behavior OR behavior change OR behavior modification OR health promotion OR health education OR preventative health care OR preventative medicine OR behavioral medicine OR behavioral health OR health-related behavior OR lifestyle change OR intervention OR medical informatics OR mHealth]*. An asterisk next to a term denotes truncation (search all terms that have this root). Different spellings of behavior (eg, *behaviour*) were accounted for in the search settings.

Study Selection

The following inclusion criteria was used for study selection: (1) articles were sampled from an adult population (≥18 years) and published in English in a peer-reviewed journal. (2) Articles reported a comparison with the mHealth app for sleep (eg, within-person pretest vs posttest or between-group experimental group vs control group), and a null hypothesis test was conducted to see if there was a statistically significant difference between comparison groups. (3) Articles reported the effects of the mHealth app intervention on a measured sleep outcome. (4) The primary intervention tool was a fully automated mHealth app for sleep accessible from a smartphone. (5) The article reported using at least one BCT in its description of the mHealth app intervention for sleep.

Article Coding Procedure

Behavior Change Techniques

All articles were independently coded by author ACA. The presence or absence of each BCT was coded to understand the frequency and use of BCTs across digital interventions for sleep (RQ1 and RQ2). The *BCT Taxonomy version 1* (BCTTv1) coding manual [12] and web-based training materials [37] were used to code for the presence or absence of BCTs. The coding manual contains labels, definitions, descriptions, and examples of each BCT category [12]. Before coding the articles, the coder ACA completed the web-based *BCTTv1* training through the official website [38]. As is common practice [39], a BCT was only coded if it was explicitly stated, if it was applied to the target behavior (ie, sleep improvement through various behaviors), and if its purpose in the mHealth app intervention was to change behavior (ie, not solely for data collection such as prompts or reminders to fill in a survey). If an article stated an external source should be retrieved for more information about components of their mHealth app for sleep, the source was followed and coded for BCTs relevant to the mHealth app used in the original study—these external sources often included preregistered protocols or electronic appendices with screenshots of the app. Frequency of a BCT's use in an intervention was not coded per coding instructions from the *BCTTv1* starter pack manual [39] and also because tracking how often a user engages with features of an app would require a hands-on review study design.

Sleep Outcomes

Overview

Sleep outcomes were only coded if they were a target of the mHealth app intervention for sleep. For example, some articles included outcomes like sleep apnea to measure prevalence in the sample and generalize to a population but not as a target of the mHealth app intervention. In cases such as these, sleep apnea was not coded.

Sleep Outcome Effect Coding

To answer RQs 1-3, the total number of positive, negative, and null sleep outcomes reported for each intervention was coded. *Positive* referred to desirable or advantageous improvements in a sleep outcome, whereas *negative* referred to undesirable or harmful effects on sleep. For example, a significant decrease in wake after sleep onset, although inherently reported as a negative number, was coded as *positive* because reducing the cumulative time spent awake after initially falling asleep is an improvement for sleep health. *Null* was coded when an intervention had a nonsignificant effect on a sleep outcome ($P>.05$). Within-person change was assessed using baseline and postintervention scores of a sleep measure. If a study reported using 2 nights of objective data at baseline to account for *first night effects* [40], the second night was used as the baseline measure.

As an exploratory step to supplement interpretation of these effects, the size of the effect of the intervention on a sleep outcome was also coded as a Cohen's *d*. This statistic is recommended for use when an outcome variable is measured in different ways (ie, different scales used to measure sleep outcomes) [41]. The mean, SD, and sample size at baseline and posttest for each sleep outcome was used to calculate the effect size Cohen's *d* using standard formulas [42,43]. Although there were not enough articles to properly pool and test for moderators (as would be done in a meta-analysis), the addition of effect sizes was included to supplement understanding of the effect of digital interventions on sleep outcomes in this study. More on the extraction and calculation of effect size data is provided in Multimedia Appendix 2.

Sleep Outcome Operationalization Coding

To answer RQ3, 2 sets of codes were applied for all sleep outcomes. First, the dimensions of sleep—sleep quality, sleep duration, sleep continuity, sleep timing, and sleepiness—that the sleep outcome measure assessed. Second, the method of measurement—subjective or objective—that was used to capture the sleep outcome. A detailed overview of the reporting of sleep dimensions and methods of measurement across all studies is provided in Table 1.

Table 1. Dimensions of sleep, definitions, and measurement.

Sleep dimension	Definition [23]	Measurement	
		Subjective	Objective
Sleep quality	Satisfaction with sleep; subjective assessment of sleep as *good* or *poor*.	Insomnia Severity Index, Pittsburg Sleep Quality Index, Sleep Diary, Sleep Condition Indicator, and Jenkins Sleep Scale	—[a]
Sleep duration	The total amount of sleep acquired in a 24-hour period.	Sleep Diary	Actigraphy, WatchPat
Sleep continuity	The ease of falling asleep and returning to sleep (sleep efficiency, wake after sleep onset, sleep onset latency, and number of awakenings).	Sleep Diary	Actigraphy, WatchPat
Sleep timing	The placement or positioning of sleep in a 24-hour period.	Sleep Timing Questionnaire and Sleep Hygiene Index	—
Sleepiness or alertness	The facility to maintain or sustain attentive wakefulness.	Epworth Sleepiness Scale, Work Productivity and Impairment, Functional Outcomes of Sleep, Glasgow Sleep Impact Index	—

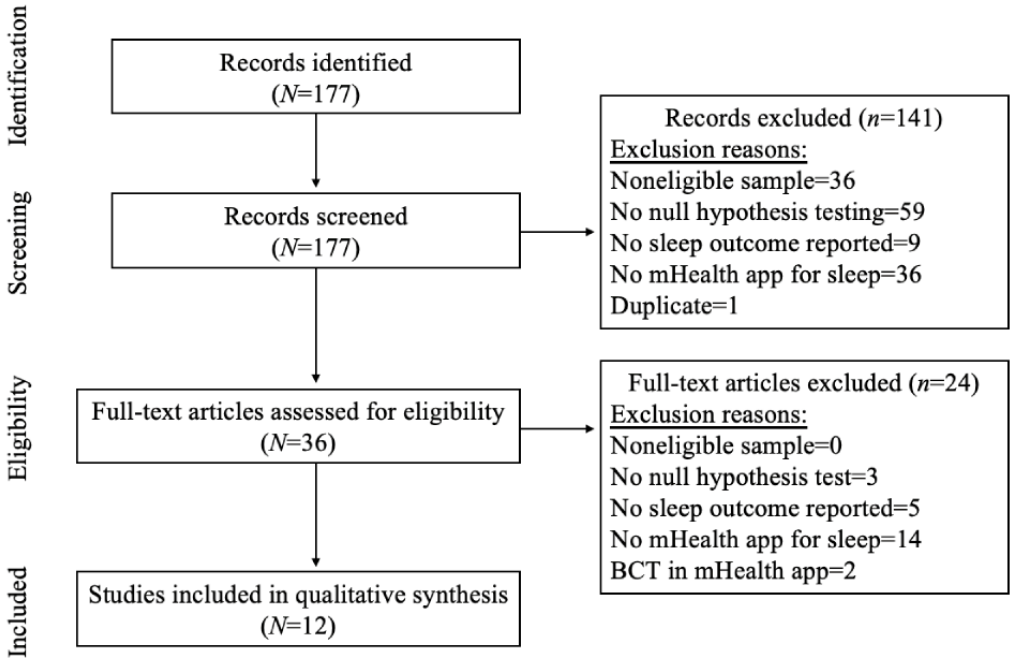
[a]Not available.

Results

Literature Search

A total of 177 articles were identified from the PubMed database search conducted on January 9, 2020. Of the 177 articles, 141 (79.7%) were excluded in the title and abstract screening. Of the 36 articles that underwent full-text review, 24 (66.7%) were excluded, resulting in 12 eligible articles in this study. Figure 1 shows a detailed overview of article screening decisions in PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) format [44].

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) reporting of article screening decisions. BCT: behavior change technique; mHealth: mobile health.



Characteristics of Included Studies

All study characteristics are reported in Tables 2 and 3. Studies were distributed evenly among those using a randomized controlled trial (RCT) design (6/12, 50%) and those using a within-person pretest–posttest design (6/12, 50%). All studies (12/12, 100%) were conducted in developed countries: 42% (5/12) in the United States; 17% (2/12) each in the Netherlands and Australia; and 8% (1/12) each in Korea, England, and Taiwan. Most participants were middle aged (mean 39; range 22-48 years), 67% female (8/12, 67%, studies had a majority female sample), and all studies that reported ethnicity had a majority of White participants in their samples (although the majority of studies did not report ethnicity). The range of intervention duration was 2 to 26 weeks, although most interventions had an average duration of 6 to 7 weeks. Of the 12 studies, 8 (67%) included a validated measure of insomnia and 4 (33%) included measures of sleep quality or sleep condition in general. Of the 8 studies with insomnia measures,

6 (75%) had samples with clinically significant levels of insomnia and 2 (25%) had samples with subthreshold insomnia.

In preliminary analyses, we examined the effects of mHealth app interventions on sleep outcomes. Across 67 coded effects, 39 (58%) sleep outcomes were positive, 28 (42%) were null, and none (0%) were negative. Across the 12 studies, 4 (33%) reported only positive findings, 7 (58%) reported mixed findings (both positive and null), and 1 (8%) reported only null findings. There were noticeable differences in the reporting of positive sleep outcomes between RCTs and pretest–posttest designs. Specifically, 47% (18/38) of sleep outcomes reported in RCTs were positive compared with 72% (21/29) of sleep outcomes reported in pretest–posttests being positive. Of the studies reporting multiple effects, a majority of pretest–posttest studies reported only positive results (4/6, 67% pretest–posttest studies reported only positive effects with no null), whereas none (0/6, 0%) of the RCT studies reported only positive results. The majority (5/6, 83%) of RCTs reported both positive and null effects.

Next, we examined if baseline insomnia of participants played a role in the outcomes reported by interventions. Of the 8 samples with baseline measures of insomnia for their participants, 6 (75%) had clinically significant insomnia and 2 (25%) had subthreshold insomnia levels. No noticeable difference in the effect of mHealth app interventions on sleep between participants with clinical insomnia and subclinical insomnia was found. Specifically, of the 6 samples, 4 (67%) samples with clinical insomnia reported more positive sleep outcomes than null for their intervention, 1 (17%) reported an even ratio of positive to null sleep outcomes, and 1 (17%) reported more null compared with positive sleep outcomes. Of the 2 samples with subthreshold insomnia, 1 (50%) reported only positive sleep outcomes whereas the other (1/2, 50%) reported a nearly even number of positive and null sleep outcomes (6 positive and 7 null). Thus, there was not clear evidence that samples with varying levels of insomnia experienced fewer positive effects of the mHealth app interventions on their sleep.

Table 2. Characteristics of included studies (N=12; part 1).

Study	Design	Country[a]	Sample size[b]	Age (years), mean (SD)[b]	Female participants, n (%)[b]	Ethnicity of participants, n (%)[b]
Horsch et al [45]	RCT[c]	The Netherlands	151	40 (13)	94 (62)	Not reported
Pulantara et al [46]	Pre-post	United States	27	36 (10)	3 (11)	White: 20 (74)
Murawski et al [24]	RCT	Australia	160	42 (10)	128 (80)	White: 146 (91) and Asian: 10 (6)
Kang et al [47]	Pre-post	Korea	19	45 (10)	12 (63)	Not reported
Espie et al [48]	RCT	English-speaking	1711	48 (14)	1329 (78)	White: 1558 (91); mixed: 36 (2); other: 35 (2); Asian: 45 (3); and Black: 19 (1)
Reilly et al [49]	Pre-post	United States	38	44 (11)	6 (16)	White: 34 (90); Black: 2 (5); Puerto Rican: 2 (5); Filipino: 2 (5); American Indian: 1 (3); and other: 2 (5)
Luik et al [50]	Pre-post	England	98	44 (15)	65 (66)	Not reported
Oftedal et al [51]	RCT	Australia	40	36 (10)	21 (53)	White: 36 (90); Asian: 2 (5); Middle Eastern: 1 (2.5); and Aboriginal or Torres Straight or Pacific Islander: 1 (2.5)
van Drongelen et al [52]	RCT	The Netherlands	502	41 (8)	34 (7)	Not reported
Bostock et al [53]	RCT	United States	270	34 (6)	90 (33)	Not reported
Horsch et al [54]	Pre-post	The Netherlands	45	35 (14)	30 (67)	Not reported
Chu et al [55]	Pre-post	Taiwan	18	22 (1)	15 (83)	Not reported

[a]If country is not specified (ie, recruited via the web without a country requirement) then the language requirement is stated.

[b]Sample size and sample characteristics are at baseline (or posttest, if unavailable). Unaccounted percentages are nonresponders. Percentages over 100% represent biracial or multiple categories selected or attributed to rounding.

[c]RCT: randomized controlled trial.

Table 3. Characteristics of included studies (N=12; part 2).

Study	Duration in weeks	Baseline insomnia (ISI[a] score)	Supplemental mode of delivery	Nonsleep intervention target	BCTs[b] used in app
Horsch et al [45]	6-7	Insomnia (16.4)	None	None	1, 2, 3, 4, 7, 8, 9, 10, 11, 12, and 13
Pulantara et al [46]	4-6	Insomnia (15.6)	Human, in-person, phone call, SMS	None	1, 2, 3, 4, 7, 8, 9, and 11
Murawski et al [24]	12	No insomnia (12.4)	Email, mail, SMS	Physical activity	1, 2, 3, 4, 5, 7, 8, 9, and 12
Kang et al [47]	4	Insomnia (20.4)	Human, in-person, phone call	None	1, 2, 4, 6, 7, 8, 9, 11, and 12
Espie et al [48]	8	N/A[c]	Website	None	1, 2, 3, 4, 7, 8, 9, 11, and 12, 13
Reilly et al [49]	6	Insomnia (16.0)	App	None	1, 2, 4, 7, 8, 11, 12, and 13
Luik et al [50]	6 20-min sessions	Insomnia (18.5)	Human, phone call, website	None	1, 2, 3, 4, 7, 8, 9, 11, 12, and 13
Oftedal et al [51]	4	N/A	Email	Physical activity, diet	1, 2, 3, 4, 5, 9, and 12
van Drongelen et al [52]	26	N/A	Website	Physical activity, nutrition	4, 11, and 12
Bostock et al [53]	8	N/A	Website	None	1, 2, 3, and 9
Horsch et al [54]	3	No insomnia (13.5)	None	None	2, 4, 7, 8, 11, and 12
Chu et al [55]	2	Insomnia (18.5)	None	None	1, 2, 4, 7, 8, 11, and 12

[a]ISI: Insomnia Severity Index. Insomnia was defined using the ISI (≥15 indicates clinical insomnia) because of its reliability and validity to detect clinical cases of insomnia [56]. Scale range 0-28. Range 0-7, no clinically significant insomnia; range 8-14, subthreshold insomnia; range 15-21, clinical insomnia (moderate); and range 22-28, clinical insomnia (severe).

[b]BCT: behavior change technique. BCT numbers (column 5) are as follows: 1, goals and planning; 2, feedback and monitoring; 3, social support; 4, shaping knowledge; 5, natural consequences; 6, comparison of behavior; 7, associations; 8, repetition and substitution; 9, comparison of outcomes; 10, reward and threat; 11, regulation; 12, antecedents; 13, identity; 14, scheduled consequences; 15, self-belief; and 16, covert learning.

[c]N/A: not applicable. Studies that did not report a validated measure of insomnia for their treatment group.

Finally, when exploring effect sizes, the average sleep outcome effect size (Cohen's *d*) across digital interventions was 0.87 (range 0.04-2.88). However, of the 67 coded effects for sleep outcomes, only 39 (58%) had sufficient information to calculate a Cohen's *d* effect size (ie, mean, SD, and sample size at baseline and after the test). This severely limited our ability to provide an additional method of interpretation for the effect of digital interventions on sleep outcomes using effect sizes. Specifically, of the 12 studies, 2 (17%) were excluded because they did not report enough information to calculate an effect size for any of their sleep outcomes. Of the remaining 10 studies, we were still unable to calculate effect sizes for 42% of the reported sleep outcomes accounted for in the positive or null outcome coding. Therefore, this study uses the primary system of sleep outcome interpretation (positive or null outcome coding) for the *Results* section and provides further detail about the effect sizes in Multimedia Appendix 2.

RQ1: The Number of BCTs Used in mHealth Apps for Sleep

RQ1 examined how many BCTs were typically used in studies and how participants slept in those studies. The average number of different BCTs used across interventions (N=12) was 7.67 of 16 (SD 2.32; range 11-3). Most interventions implemented several different BCTs, with 75% (9/12) of studies reporting using ≥7 BCTs in their mHealth app intervention for sleep. Increasing or decreasing the number of BCTs did not seem to produce a discernible pattern in the proportion of positive sleep outcomes (Table 4). As one way to compare, we grouped studies that used ≥8 or ≤7 BCTs, creating the split point at the average number of BCTs used in all studies. The average percentage of positive outcomes for studies using ≥8 BCTs was 66.7% positive (SD 26.56%) and the average for studies using ≤7 BCTs was 60% positive (SD 45.41%), indicating a negligible difference between the 2 groups. Moreover, both the groups (≥8 and ≤7 BCTs) had very high SDs, indicating a large within-group variability. For example, the 2 studies using 7 BCTs had low congruence, where 1 study had 0% positive outcomes and the other had 100% positive outcomes. Thus, this suggests that simply increasing or decreasing the total number of BCTs is not automatically related to more positive sleep outcomes.

Table 4. Number of behavior change techniques (BCTs) used in an intervention and sleep outcomes reported in those interventions (N=12).

Study	Number of BCTs	Sleep outcomes	
		Number of positive outcomes, n (%)[a]	Number of null outcomes, n (%)
Horsch et al [45]	11	5 (50)	5 (50)
Espie et al [48]	10	3 (60)	2 (40)
Luik et al [50]	10	1 (100)	0 (0)
Murawski et al [24]	9	6 (46)	7 (54)
Kang et al [47]	9	7 (78)	2 (22)
Pulantara et al [46]	8	4 (100)	0 (0)
Reilly et al [49]	8	3 (33)	6 (67)
Oftedal et al [51]	7	0 (0)	2 (100)
Chu et al [55]	7	5 (100)	0 (0)
Horsch et al [54]	6	1 (100)	0 (0)
Bostock et al [53]	4	3 (75)	1 (25)
van Drongelen et al [52]	3	1 (25)	3 (75)

[a]The percentage of positive outcomes for sleep was calculated using the number of outcomes that were positive divided by the total number of outcomes (sum of the number of positive results and number of null results) reported across studies using this number of BCTs.

RQ2: The Type of BCTs Used in mHealth Apps for Sleep

RQ2 examined the use of specific BCTs across interventions and whether their presence was important for sleep outcomes. As reported in Table 5, the BCTs that appeared most often across mHealth app interventions for sleep were feedback and monitoring, and shaping knowledge. Other BCTs that were frequently implemented by most (≥75%) of the interventions were goals and planning, antecedents, associations, repetition and substitution, and regulation. Conversely, some BCTs were rarely or never used: natural consequences, comparison of behavior, reward and threat, scheduled consequences, self-belief, and covert learning.

As a consequence of the frequent use of the same BCTs across studies, we were unable to examine unique effects of BCTs on sleep outcomes. This collinearity among interventions using the same BCTs made it impossible to discern a pattern of positive or null outcomes associated with the use of specific BCTs in interventions. Instead, we examined if there was a presence effect: if the presence of specific BCTs in an intervention was associated with a greater percentage of positive sleep outcomes reported (Table 5). The BCT comparison of behavior stood out because when it was included in interventions the rate of positive sleep outcomes was higher (78%) than the rate of positive sleep outcomes for other BCTs (that seemed to center around 50%-60%). Another BCT that stood out was natural consequences because when it was included in studies, the rate of positive sleep outcomes was low (40%) compared with the others. However, it is important to note that the BCT comparison of behavior only had 1 study that used it and the BCT natural consequences only had 2 studies, and thus, we would need more evidence to know if this is a reliable effect of the BCT or random chance.

Table 5. Frequency of behavior change techniques (BCTs) used in mobile health app interventions across all studies (N=12).

BCT	Studies using BCT, n (%)	Positive outcomes, n (%)[a]	Total number of outcomes
Feedback and monitoring	11 (92)	38 (60)	63
Shaping knowledge	11 (92)	36 (57)	63
Goals and planning	10 (83)	37 (60)	62
Antecedents	10 (83)	32 (54)	59
Associations	9 (75)	35 (61)	57
Repetition and substitution	9 (75)	35 (61)	57
Regulation	9 (75)	30 (63)	48
Comparison of outcomes	8 (67)	29 (60)	48
Social support	7 (58)	22 (56)	39
Identity	4 (33)	12 (48)	25
Natural consequences	2 (17)	6 (40)	15
Comparison of behavior	1 (8)	7 (78)	9
Reward and threat	1 (8)	5 (50)	5
Scheduled consequences	0 (0)	—[b]	—
Self-belief	0 (0)	—	—
Covert learning	0 (0)	—	—

[a]The percentage of positive outcomes was calculated using the total number of outcomes that were positive divided by the total number of outcomes reported across all studies using this specific BCT.

[b]None of the studies used this BCT.

RQ3: Results by Dimension and Measure of Sleep

To further our interpretation of how mHealth app interventions influence sleep, the effect that interventions had on sleep was further broken down by dimension. Sleep quality was by far the most frequently targeted dimension of sleep with all studies including a measure for this dimension. Sleep quality also had the most reliable improvement compared with other dimensions of sleep (n=24, 73% of the 33 sleep quality outcomes were positive). The remaining dimensions of sleep were infrequently targeted by interventions (less than half of the interventions reported them). The sleep dimension that was targeted least across studies was sleep timing (2/12, 17%, studies), followed by sleep duration (3/12, 25%, studies), sleep continuity (3/12, 25%, studies), and sleepiness (5/12, 42%, studies). These dimensions also had a lower rate of positivity compared with sleep quality. Specifically, 73% (24/33) of sleep quality outcomes were positive compared with 0% (0/3) of sleep duration, 50% (2/4) of sleep timing, 60% (6/10) of sleepiness, and 67% of sleep continuity. The disproportionate appearance of sleep quality (33/67, 49%, sleep outcomes) compared with the other sleep dimensions (range of 3-10 outcomes per dimension) render these differences speculative and more studies with greater representation of dimensions would be needed to make fair comparisons between dimensions.

Beyond dimension, we examined whether the effect of mHealth app interventions on sleep differed based on whether sleep was subjectively or objectively measured. All the 12 studies used at least one subjective measure of sleep. Surprisingly, although 42% (5/12) of studies mentioned using objective measures of sleep within their intervention, only 17% (2/12) of studies reported results for objectively measured sleep. This brings into question why some studies did not report all the sleep outcomes they collected. Here, we provide a description of how sleep was measured across studies, but we are unable to make meaningful interpretations because of limited reporting of objective sleep measures. The 2 studies that reported objective measures of sleep had an average sleep outcome positivity rate of 78% and 33%. These 2 studies were above and below the study-wide average positivity rate of 58% and therefore do not give a clear indication of whether mHealth app interventions' ability to improve sleep outcomes differs when distinguishing between subjectively and objectively measured sleep. Multimedia Appendix 3 [24,45-55] details the subjective and objective measures used for sleep outcomes across studies.

Discussion

Principal Findings

The purpose of this review was to examine the use of mHealth apps for sleep published in the peer-reviewed literature. Specifically, we aimed to report and describe the use of BCTs in mHealth apps for sleep and if their effects on sleep outcomes varied by dimension and method of measurement. We found that studies most often reported positive sleep outcomes from their interventions, with no adverse effects reported. This finding suggested that mHealth apps for sleep can have desirable effects on sleep outcomes, and results were in line with previous research supporting the efficacy of mHealth apps to improve other health outcomes such as physical activity [57], sedentary

behavior [58], and chronic disease management [59]. We also found that the most commonly measured dimension of sleep was sleep quality and that objective measures of sleep were vastly underrepresented compared with subjective measures of sleep.

Regardless of sleep dimension or measurement, there were no negative effects of interventions on sleep reported across all studies. This was surprising considering research on orthosomnia indicating some adverse effects of sleep apps [8,60]. Our results could mean that reports of orthosomnia may not be attributed to mHealth apps rather individual characteristics in a subset of the population (eg, people with severe insomnia not included in this review) or attributed to short-term effects of sleep apps that dissolve after a few uses (and therefore were not captured in the studies we reviewed). Alternatively, they could be because of measurement bias in which measures to capture adverse effects such as orthosomnia were not included in the studies we reviewed. Despite several reports of orthosomnia caused by mHealth app interventions [8,60], to our knowledge, scientists are yet to develop a measure capturing this adverse effect—meaning adverse effects could be happening but are not being captured. The lack of negative findings could also be because of bias in reporting attributed to selective reporting of sleep outcomes only showing positive effects. As is discussed in detail in the *Implications for Sleep Outcomes* section, there was extreme variability in the reporting of sleep scales and subscales across studies, which impeded our ability to fully assess subcomponents of RQ1 and RQ2.

RQ1 examined if there was a pattern in the number of BCTs used in mHealth app interventions for sleep. We found that most interventions used several different BCTs in their apps instead of focusing on only a few. This follows from the previous work on mHealth targeting other health outcomes demonstrating a wide variety of BCTs implemented [59,61]. However, we were unable to determine if using a wide variety of BCTs in sleep apps was advantageous for sleep outcomes. Having several BCTs in an intervention could increase the likelihood of at least one BCT being able to help an individual improve their sleep, or could also lead to a disjointed and unpredictable experience with the app and potentially result in adverse associations with sleep. A meta-analysis on eHealth interventions for alcohol consumption did suggest it is better to focus on specific, rather than several, BCTs [62], and this could hold true for sleep apps. However, whether there is an ideal number of BCTs to implement or if it varies between dimensions of sleep is unknown and should be a major focus of future research.

Beyond examining the number of BCTs, RQ2 examined whether there were specific BCTs used more or less often across all mHealth apps for sleep. Although there was overlap in the BCTs deployed across studies, there were some BCTs that noticeably appeared more often than others across interventions. The BCTs feedback and monitoring and goal setting are some of the most commonly implemented BCTs across mHealth app interventions targeting physical activity and diet [63], and this held in our review of mHealth apps for sleep. Furthermore, the BCTs that appeared in almost all the interventions included: shaping knowledge, antecedents, associations, repetition and substitution, and regulation. The widespread use of these BCTs supports the hypotheses that techniques focused on aspects of behavior before sleep (eg, education about antecedents, habit formation, reducing negative emotion) may be more suitable in mHealth for sleep than BCTs focusing on future outcomes or consequences of sleep.

Indeed, we found most of the BCTs omitted or rarely used tended to focus on outcomes or consequences of sleep, and included scheduled consequences, reward and threat, natural consequences, and covert learning. This could reflect an understanding that sleep differs from other health outcomes in that it is not always under one's control and that it may be better for sleep apps to focus on predictors of sleep rather than the outcomes or consequences of it. The omission of comparison of behavior suggests that despite being a major component in other mHealth app interventions [64], social components may not be as relevant in mHealth for sleep because of its innately solitary nature. The omission of self-belief was initially surprising considering it does not focus on social components or outcomes of sleep. However, social cognitive theory [65] suggests that self-efficacy, similar to self-belief, works in tandem with outcome expectations (part of natural consequences), so it may not make sense to target one without the other [66]. It is therefore understandable why self-belief was not promoted because of its relationship with outcome expectancies (part of natural consequences). Although we were unable to assess if these BCTs were associated with worse outcomes, future research could conduct optimization trials [67] to understand which BCTs, or combinations of BCTs, are effective.

RQ2 also examined if the presence of specific BCTs was associated with a larger percentage of positive sleep outcomes reported. Although the conclusions we can draw are limited because of the small number of studies, there were 2 BCTs that stood out. The BCT comparison of behavior was used in studies with the highest proportion of positive results, whereas natural consequences was used in studies with the lowest proportion of positive results. Both findings would map on to hypotheses that focusing on aspects of behavior related to sleep may be more advantageous than focusing on consequences of sleep.

The first part of RQ3 examined the effect of mHealth app interventions by sleep dimension. Sleep quality was the most consistently targeted and improved dimension of sleep by mHealth apps. This finding is encouraging because previous research suggests improvements in sleep quality is the most important indicator of the restorative benefits of sleep [24,68]. This may partly explain why mHealth apps seem to be so good at improving sleep quality. Our results also bring into question why other dimensions of sleep such as sleep duration do not share the same reliable improvements as sleep quality, especially as sleep duration is the key indicator of sleep health used by medical professionals [69]. Our findings warrant future examination of whether all dimensions of sleep can be improved by mHealth apps or whether mHealth app efficacy varies by sleep dimension.

The second part of RQ3 examined whether the effect of mHealth app interventions was consistent across subjectively and objectively measured sleep outcomes. This was important to

assess because of differences in the long-term health outcomes associated with subjective and objective sleep [70,71], and their potential to differentially predict treatment efficacy in insomnia intervention research [29]. Unfortunately, we were unable to answer this question because of limited information provided from articles (ie, less than half of the interventions using objective measures reported them in their results). A similar issue was noted in a recent systematic review of smartphone-delivered interventions for health behaviors [72]. This problem highlights an overarching issue with the reporting of sleep outcomes that made it difficult to assess and compare sleep outcomes across interventions.

Implications for Sleep Outcomes

When testing our RQs, we came across issues related to sleep outcome reporting and interpretation. Many studies did not use thresholds that are significant based on clinical research and instead relied on statistical significance as an indication for whether a sleep outcome improved and the intervention *worked*. This is problematic because the statistical improvement of a sleep outcome does not equate to a clinically meaningful improvement. The former would not be considered effective if it were used by clinicians in a sleep practice. There are excellent thresholds for clinically meaningful change laid out by [22] (eg, cumulative time spent awake after initially falling asleep should be <30 minutes); however, they were not referenced in the interpretation of sleep data. Furthermore, there were sometimes misleading conclusions about sleep improvement because outcomes were assessed individually instead of in relation to other relevant outcomes. For example, a significant increase in total sleep time by itself appears desirable, but if sleep efficiency (total sleep time/time in bed) decreased or did not improve, this could be an adverse effect of the intervention as it could indicate individuals spent more time in bed restless.

There are models for how to use a more systematic and comprehensive approach to sleep measurement and reporting. For example, the work by Carney et al [73] was meant to facilitate comparison of sleep across studies by creating a standardized sleep diary, a popular measure in sleep research. Their work resulted in 9 items for the standard sleep diary that are used to calculate 8 critical sleep indices [73]. Despite this consensus in 2012, neither of the studies that used a sleep diary in this review reported all 8 indices. Unfortunately, this was not an isolated issue to sleep diaries, and selective reporting continued while examining other measures of sleep. There was extreme variability in reporting of scales' global scores, subscales, and items both within and across studies using the same scales (Multimedia Appendix 3). For instance, 7 studies used the Pittsburg Sleep Quality Index, but 4 studies reported only the global score, 1 study reported the global and all subscales, 1 study reported the global and 1 of 7 subscales, and 1 study reported no global but just 3 of 7 subscales. Although most studies using the Insomnia Severity Index used the global score derived from its 7 items, 1 study derived a global score from just 5 of the 7 items and also reported 4 of the 5 collected items individually. This lack of consensus is problematic as it makes synthesizing findings across these studies while controlling for the type of outcome measure nearly impossible, thereby precluding approaches such as meta-analysis that are needed to inform medical decision-making.

The inconsistent reporting of sleep outcomes is concerning for many reasons, including that there was no explanation for why some metrics (subscales or individual items) were highlighted, whereas others were omitted. The selective reporting also brought into question issues related to reporting bias (eg, if the omitted sleep metrics did not support hypotheses). The finding that none of the RCTs reported only positive results whereas most pretest–posttest studies reported only positive results could support this point given most RCTs have preregistration, which deters selective reporting of outcomes. Although RCTs provide the highest level of evidence to make causal inferences [74], pretest–posttest designs have a benefit of requiring fewer resources to execute, although they also have drawbacks including their inability to control for third variables because of a lack of randomization. The difference in positive sleep outcome reporting by study design could suggest that pretest-posttest studies are inflating the effect of digital interventions on sleep and that RCTs present a more accurate and variable picture of the potential for mHealth app interventions to improve sleep. It could also be because of random chance—our sample of studies was relatively small and it is possible that a larger sample would neutralize this pattern and find no difference in reporting by study design. To answer this question, it is important to address selective reporting of sleep outcomes as a field, potentially through an increased requirement for preregistration, even for pretest-posttest study designs.

The selective reporting of sleep outcomes presents challenges for the field when trying to determine the effect sizes of mHealth app interventions on different domains of sleep across studies. Whether there were valid reasons for their omission was not explained, but regardless of reasoning the field would benefit from reporting all collected sleep metrics to increase transparency and examination across studies (ie, all possible metrics of recorded sleep data are reported somewhere like appendices). Sleep researchers could use work by Buysse [23], Buysse et al [75], and Morin [22] to create an ontology and taxonomy for consensus on sleep measurement and reporting. This would create a shared language and expedite communication of information and results so researchers are not advancing in silos but rather the field is advancing together with increased speed and efficiency. It is also of critical importance that these conversations include clinicians and researchers to promote the collaboration across domains through similarly defined variables of sleep and the importance designated to them.

Limitations and Future Directions

Although this study had several strengths including being the first (to our knowledge) to systematically examine the use of BCTs in peer-reviewed studies of fully automated mHealth apps for sleep, there were notable limitations to our study. A limitation inherent to all systematic reviews is its equal treatment of studies regardless of sample and effect sizes. We tried to account for this by including the effect sizes of sleep outcomes when possible; however, limited data and overall consensus in

reporting precluded a full understanding of the overall effect. A meta-analysis could provide a more complete picture because it weighs studies according to the effect and sample size and can detect heterogeneity in effect sizes to identify subgroups of people for whom the intervention is more or less effective (an approach that is critical to the increasingly employed precision medicine approach). In addition, we were unable to model the co-occurrence of BCTs used in interventions which could be hiding negative effects (eg, if one BCT is adverse and the other is beneficial, their effect would be null) or a potential synergistic effect where combinations of BCTs are more effective together than when used alone.

Similarly, it was not possible to test if co-occurring targets of the intervention (eg, promoting physical activity in conjunction with sleep) could moderate the efficacy of the intervention on sleep outcomes. It has been suggested that targeting multiple health behaviors together can lead to greater health improvements than targeting one alone [76] because of spillover effects (eg, transfer or gateway effects) in which success with one health behavior aids in the ability to succeed with a second health behavior [77]. This may be relevant for behaviors such as physical activity and sleep as they are known to have a reciprocal relationship with one another [78,79]. By contrast, targeting multiple health behaviors at once could fail to address either behavior in sufficient depth, thereby reducing the intervention's potential to be effective [80-82]. This would be in line with theory that suggests addressing multiple health behaviors requires significant effort (cross-behavior regulation [83]) and that the effort put toward improving one health behavior could limit one's ability to improve another (self-control strength model and ego depletion [84,85]).

Another potential confound we were unable to account for was the multi-model intervention delivery involving a human coach. Of the 12 studies, 3 (25%) had a live human component that may have enhanced overall effectiveness of the intervention on sleep outcomes. Indeed, there is some evidence to suggest that human guidance of internet-based interventions can improve intervention efficacy [86] and also promote engagement with the digital intervention [87]. Although there are benefits to adding human guidance, there are also potential drawbacks, including the addition of human support increasing cost of the intervention and burden on health care providers. Important factors to examine in regard to the addition of human support in digital interventions are whether the quantity of human support, quality of human support (ie, level of expertise), frequency, timing, or mode of delivery (in-person, via phone call, or video chat) matter for the overall efficacy of digital interventions on sleep.

Although it was not a focus of this study, an important next step is to examine the use of sleep hygiene in conjunction with BCTs. Sleep hygiene is a set of behavioral and environmental recommendations to promote sleep [88] and is often the first line of defense and treatment for sleep disorders [64,89,90]. Sleep hygiene is known to improve sleep in clinical and nonclinical populations [64], and it has been noted that popular commercial sleep apps are well-equipped to support sustainable sleep hygiene practices [10]. Improvements in sleep hygiene behaviors mediated sleep quality in a recent mHealth app intervention study [91] in which 30% of the changes in sleep quality were explained by changes in sleep hygiene. A natural extension and future direction of our study is adding a meta-analysis to examine efficacy while weighing according to sample sizes and identify the role of moderators of any observed effects (eg, sleep hygiene, co-occurring BCTs, or cotargeted health behaviors).

Conclusions

This study conducted a systematic review of published peer-reviewed articles from 2010 to 2020 on mHealth app interventions for sleep, their use of BCTs, and their effect on sleep outcomes. Overall, we found overwhelming evidence that sleep apps can be effective at improving sleep, and we did not come across any reports of adverse effects (orthosomnia). However, this does not mean that adverse effects did not occur, and we recommend future research work on revising standards of sleep outcome measurement and reporting. We found evidence that the type of BCTs used in mHealth apps for sleep differed from other health outcomes, which suggests mHealth app intervention components may not be a one-size-fits-all and that sleep apps may require different design from other health app interventions. Further research is needed with improved measures and reporting of sleep to identify the optimal design components and potential limitations of mHealth app interventions for sleep.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Behavior change techniques and examples with sleep.
[DOCX File , 23 KB - mhealth_v10i4e33527_app1.docx]

Multimedia Appendix 2
Cohen's *d* effect size interpretations.
[DOCX File , 20 KB - mhealth_v10i4e33527_app2.docx]

Multimedia Appendix 3

Measures used for sleep outcomes across studies.
[DOCX File , 42 KB - mhealth_v10i4e33527_app3.docx]

References

1. Duncan M, Murawski B, Short CE, Rebar AL, Schoeppe S, Alley S, et al. Activity trackers implement different behavior change techniques for activity, sleep, and sedentary behaviors. Interact J Med Res 2017 Aug 14;6(2):e13 [FREE Full text] [doi: 10.2196/ijmr.6685] [Medline: 28807889]
2. Marcolino MS, Oliveira JA, D'Agostino M, Ribeiro AL, Alkmim MB, Novillo-Ortiz D. The impact of mhealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
3. Williams SL, French DP. What are the most effective intervention techniques for changing physical activity self-efficacy and physical activity behaviour--and are they the same? Health Educ Res 2011 Apr;26(2):308-322 [FREE Full text] [doi: 10.1093/her/cyr005] [Medline: 21321008]
4. mHealth: new horizons for health through mobile technologies. World Health Organization. 2011. URL: https://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2020-08-13]
5. National center for health statistics: leading causes of death. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm [accessed 2020-09-22]
6. Walker M. Why We Sleep: Unlocking the Power of Sleep and Dreams. New York: Simon and Schuster; 2017.
7. Mead MP, Irish LA. Application of health behaviour theory to sleep health improvement. J Sleep Res 2020 Oct 22;29(5):e12950 [FREE Full text] [doi: 10.1111/jsr.12950] [Medline: 31758596]
8. Baron KG, Abbott S, Jao N, Manalo N, Mullen R. Orthosomnia: are some patients taking the quantified self too far? J Clin Sleep Med 2017 Mar 15;13(2):351-354 [FREE Full text] [doi: 10.5664/jcsm.6472] [Medline: 27855740]
9. Aji M, Gordon C, Stratton E, Calvo RA, Bartlett D, Grunstein R, et al. Framework for the design engineering and clinical implementation and evaluation of mHealth apps for sleep disturbance: systematic review. J Med Internet Res 2021 Feb 17;23(2):e24607 [FREE Full text] [doi: 10.2196/24607] [Medline: 33595441]
10. Grigsby-Toussaint DS, Shin JC, Reeves DM, Beattie A, Auguste E, Jean-Louis G. Sleep apps and behavioral constructs: a content analysis. Prev Med Rep 2017 Jun;6:126-129 [FREE Full text] [doi: 10.1016/j.pmedr.2017.02.018] [Medline: 28316907]
11. Peake JM, Kerr G, Sullivan JP. A critical review of consumer wearables, mobile applications, and equipment for providing biofeedback, monitoring stress, and sleep in physically active populations. Front Physiol 2018;9:743 [FREE Full text] [doi: 10.3389/fphys.2018.00743] [Medline: 30002629]
12. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
13. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
14. Michie S, Fixsen D, Grimshaw JM, Eccles MP. Specifying and reporting complex behaviour change interventions: the need for a scientific method. Implement Sci 2009;4:40 [FREE Full text] [doi: 10.1186/1748-5908-4-40] [Medline: 19607700]
15. Behaviour change: individual approaches. National Institute for HealthCare Excellence, United Kingdom. 2014. URL: https://www.nice.org.uk/guidance/ph49/resources/behaviour-change-individual-approaches-pdf-1996366337989 [accessed 2021-03-15]
16. Alvaro PK, Roberts RM, Harris JK. A systematic review assessing bidirectionality between sleep disturbances, anxiety, and depression. Sleep 2013 Jul 01;36(7):1059-1068 [FREE Full text] [doi: 10.5665/sleep.2810] [Medline: 23814343]
17. Brown TA, Chorpita BF, Barlow DH. Structural relationships among dimensions of the DSM-IV anxiety and mood disorders and dimensions of negative affect, positive affect, and autonomic arousal. J Abnorm Psychol 1998 May;107(2):179-192. [doi: 10.1037//0021-843x.107.2.179] [Medline: 9604548]
18. Fuller KH, Waters WF, Binks PG, Anderson T. Generalized anxiety and sleep architecture: a polysomnographic investigation. Sleep 1997 May;20(5):370-376. [doi: 10.1093/sleep/20.5.370] [Medline: 9381061]
19. Regestein QR, Dambrosia J, Hallett M, Murawski B, Paine M. Daytime alertness in patients with primary insomnia. Am J Psychiatry 1993 Oct;150(10):1529-1534. [doi: 10.1176/ajp.150.10.1529] [Medline: 8379559]
20. Staner L. Sleep and anxiety disorders. Dialogues Clin Neurosci 2003 Sep;5(3):249-258 [FREE Full text] [Medline: 22033804]
21. Stepanski E, Zorick F, Roehrs T, Young D, Roth T. Daytime alertness in patients with chronic insomnia compared with asymptomatic control subjects. Sleep 1988 Feb;11(1):54-60. [doi: 10.1093/sleep/11.1.54] [Medline: 3363270]
22. Morin CM. Measuring outcomes in randomized clinical trials of insomnia treatments. Sleep Med Rev 2003 Jun;7(3):263-279. [doi: 10.1053/smrv.2002.0274] [Medline: 12927124]
23. Buysse DJ. Sleep health: can we define it? Does it matter? Sleep 2014 Jan 01;37(1):9-17 [FREE Full text] [doi: 10.5665/sleep.3298] [Medline: 24470692]

24. Murawski B, Plotnikoff RC, Rayward AT, Oldmeadow C, Vandelanotte C, Brown WJ, et al. Efficacy of an m-health physical activity and sleep health intervention for adults: a randomized waitlist-controlled trial. Am J Prev Med 2019 Oct;57(4):503-514. [doi: 10.1016/j.amepre.2019.05.009] [Medline: 31542128]
25. Argyropoulos SV, Hicks JA, Nash JR, Bell CJ, Rich AS, Nutt DJ, et al. Correlation of subjective and objective sleep measurements at different stages of the treatment of depression. Psychiatry Res 2003 Sep;120(2):179-190. [doi: 10.1016/s0165-1781(03)00187-2]
26. Armitage R, Trivedi M, Hoffmann R, Rush AJ. Relationship between objective and subjective sleep measures in depressed patients and healthy controls. Depress Anxiety 1997;5(2):97-102. [doi: 10.1002/(sici)1520-6394(1997)5:2<97::aid-da6>3.0.co;2-2] [Medline: 9262940]
27. Rotenberg V, Indursky P, Kayumov L, Sirota P, Melamed Y. The relationship between subjective sleep estimation and objective sleep variables in depressed patients. Int J Psychophysiol 2000 Sep;37(3):291-297. [doi: 10.1016/s0167-8760(00)00110-0]
28. Edinger JD, Fins AI, Glenn DM, Sullivan RJ, Bastian LA, Marsh GR, et al. Insomnia and the eye of the beholder: are there clinical markers of objective sleep disturbances among adults with and without insomnia complaints? J Consult Clin Psychol 2000;68(4):586-593. [doi: 10.1037/0022-006x.68.4.586]
29. Lauderdale DS, Knutson KL, Yan LL, Liu K, Rathouz PJ. Self-reported and measured sleep duration: how similar are they? Epidemiology 2008 Nov;19(6):838-845 [FREE Full text] [doi: 10.1097/EDE.0b013e318187a7b0] [Medline: 18854708]
30. Lee JH, Reynolds CF, Hoch CC, Buysse DJ, Mazumdar S, George CJ, et al. Electroencephalographic sleep in recently remitted, elderly depressed patients in double-blind placebo-maintenance therapy. Neuropsychopharmacology 1993 Mar;8(2):143-150. [doi: 10.1038/npp.1993.16] [Medline: 8471126]
31. Zachariae R, Lyby MS, Ritterband LM, O'Toole MS. Efficacy of internet-delivered cognitive-behavioral therapy for insomnia - A systematic review and meta-analysis of randomized controlled trials. Sleep Med Rev 2016 Dec;30:1-10. [doi: 10.1016/j.smrv.2015.10.004] [Medline: 26615572]
32. Turner-McGrievy GM, Hales SB, Schoffman DE, Valafar H, Brazendale K, Weaver RG, et al. Choosing between responsive-design websites versus mobile apps for your mobile behavioral intervention: presenting four case studies. Transl Behav Med 2016 Nov 03 [FREE Full text] [doi: 10.1007/s13142-016-0448-y] [Medline: 27812798]
33. Altevogt B, Colten H. Sleep disorders and sleep deprivation: an unmet public health problem. Nat Acad Press 2006. [doi: 10.17226/11617] [Medline: 20669438]
34. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
35. The App Store turns 10. Apple Newsroom. 2018. URL: https://www.apple.com/newsroom/2018/07/app-store-turns-10/ [accessed 2020-08-13]
36. Gusenbauer M, Haddaway NR. Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. Res Synth Methods 2020 Mar;11(2):181-217 [FREE Full text] [doi: 10.1002/jrsm.1378] [Medline: 31614060]
37. UCL BCT Taxonomy - Home. UCL Centre for Behaviour Change. 2014. URL: https://www.bct-taxonomy.com/ [accessed 2020-08-13]
38. BCT Taxonomy Training. URL: https://www.bct-taxonomy.com/ [accessed 2021-09-10]
39. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
40. Toussaint M, Luthringer R, Schaltenbrand N, Carelli G, Lainey E, Jacqmin A, et al. First-night effect in normal subjects and psychiatric inpatients. Sleep 1995 Jul;18(6):463-469. [doi: 10.1093/sleep/18.6.463] [Medline: 7481418]
41. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol 2013 Nov 26;4:863 [FREE Full text] [doi: 10.3389/fpsyg.2013.00863] [Medline: 24324449]
42. Borenstein M, Hedges L, Higgins J, Rothstein H. Comprehensive meta-analysis, version 2 Biostat. Englewood NJ. 2005. URL: https://www.meta-analysis.com/ [accessed 2021-09-10]
43. Cohen J. Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Hillsdale, NJ: Psychology Press; 1988.
44. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
45. Horsch CH, Lancee J, Griffioen-Both F, Spruit S, Fitrianie S, Neerincx MA, et al. Mobile phone-delivered cognitive behavioral therapy for insomnia: a randomized waitlist controlled trial. J Med Internet Res 2017 Apr 11;19(4):e70 [FREE Full text] [doi: 10.2196/jmir.6524] [Medline: 28400355]
46. Pulantara IW, Parmanto B, Germain A. Clinical feasibility of a just-in-time adaptive intervention app (iREST) as a behavioral sleep treatment in a military population: feasibility comparative effectiveness study. J Med Internet Res 2018 Dec 07;20(12):e10124 [FREE Full text] [doi: 10.2196/10124] [Medline: 30530452]

47. Kang SG, Kang JM, Cho SJ, Ko KP, Lee YJ, Lee HJ, et al. Cognitive behavioral therapy using a mobile application synchronizable with wearable devices for insomnia treatment: a pilot study. J Clin Sleep Med 2017 Apr 15;13(4):633-640 [FREE Full text] [doi: 10.5664/jcsm.6564] [Medline: 28162145]
48. Espie CA, Emsley R, Kyle SD, Gordon C, Drake CL, Siriwardena AN, et al. Effect of digital cognitive behavioral therapy for insomnia on health, psychological well-being, and sleep-related quality of life: a randomized clinical trial. JAMA Psychiatry 2019 Jan 01;76(1):21-30. [doi: 10.1001/jamapsychiatry.2018.2745] [Medline: 30264137]
49. Reilly ED, Robinson SA, Petrakis BA, Kuhn E, Pigeon WR, Wiener RS, et al. Mobile app use for insomnia self-management: pilot findings on sleep outcomes in veterans. Interact J Med Res 2019 Jul 24;8(3):e12408 [FREE Full text] [doi: 10.2196/12408] [Medline: 31342904]
50. Luik AI, Bostock S, Chisnall L, Kyle SD, Lidbetter N, Baldwin N, et al. Treating depression and anxiety with digital cognitive behavioural therapy for insomnia: a real world NHS evaluation using standardized outcome measures. Behav Cogn Psychother 2017 Jan;45(1):91-96. [doi: 10.1017/S1352465816000369] [Medline: 27456542]
51. Oftedal S, Burrows T, Fenton S, Murawski B, Rayward AB, Duncan MJ. Feasibility and preliminary efficacy of an m-Health intervention targeting physical activity, diet, and sleep quality in shift-workers. Int J Environ Res Public Health 2019 Oct 10;16(20) [FREE Full text] [doi: 10.3390/ijerph16203810] [Medline: 31658624]
52. van Drongelen A, Boot CR, Hlobil H, Twisk JW, Smid T, van der Beek J. Evaluation of an mHealth intervention aiming to improve health-related behavior and sleep and reduce fatigue among airline pilots. Scand J Work Environ Health 2014 Nov;40(6):557-568. [doi: 10.5271/sjweh.3447] [Medline: 25121620]
53. Bostock S, Luik AI, Espie CA. Sleep and productivity benefits of digital cognitive behavioral therapy for insomnia: a randomized controlled trial conducted in the workplace environment. J Occup Environ Med 2016 Dec;58(7):683-689. [doi: 10.1097/JOM.0000000000000778] [Medline: 27257747]
54. Horsch C, Spruit S, Lancee J, van Eijk R, Beun RJ, Neerincx M, et al. Reminders make people adhere better to a self-help sleep intervention. Health Technol (Berl) 2017 Dec 23;7(2):173-188 [FREE Full text] [doi: 10.1007/s12553-016-0167-x] [Medline: 29201588]
55. Chu HC, Liu YM, Kuo FR. A mobile sleep-management learning system for improving students' sleeping habits by integrating a self-regulated learning strategy: randomized controlled trial. JMIR Mhealth Uhealth 2018 Oct 29;6(10):e11557 [FREE Full text] [doi: 10.2196/11557] [Medline: 30373729]
56. Morin CM, Belleville G, Bélanger L, Ivers H. The Insomnia Severity Index: psychometric indicators to detect insomnia cases and evaluate treatment response. Sleep 2011 May 01;34(5):601-608 [FREE Full text] [doi: 10.1093/sleep/34.5.601] [Medline: 21532953]
57. Flores-Mateo G, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile phone apps to promote weight loss and increase physical activity: a systematic review and meta-analysis. J Med Internet Res 2015;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
58. Elavsky S, Knapova L, Klocek A, Smahel D. Mobile health interventions for physical activity, sedentary behavior, and sleep in adults aged 50 years and older: a systematic literature review. J Aging Phys Act 2019 Mar 06:1-29. [doi: 10.1123/japa.2017-0410] [Medline: 30507266]
59. Garnett CV, Crane D, Brown J, Kaner EF, Beyer FR, Muirhead CR, et al. Behavior change techniques used in digital behavior change interventions to reduce excessive alcohol consumption: a meta-regression. Ann Behav Med 2018 May 18;52(6):530-543. [doi: 10.1093/abm/kax029] [Medline: 29788261]
60. Aupetit S, Dubroca G, Escaich S, Cabon P. A qualitative study of sleep trackers usage: evidence of orthosomnia. In: Contemp Ergon Hum Factors. Birmingham UK: CIEHF; 2019:7.
61. Mercer K, Li M, Giangregorio L, Burns C, Grindrod K. Behavior change techniques present in wearable activity trackers: a critical analysis. JMIR Mhealth Uhealth 2016;4(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4461] [Medline: 27122452]
62. Black N, Mullan B, Sharpe L. Computer-delivered interventions for reducing alcohol consumption: meta-analysis and meta-regression using behaviour change techniques and theory. Health Psychol Rev 2016 Sep;10(3):341-357. [doi: 10.1080/17437199.2016.1168268] [Medline: 26999311]
63. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. Ann Behav Med 2017 Apr;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
64. Yang C, Maher JP, Conroy DE. Implementation of behavior change techniques in mobile applications for physical activity. Am J Prev Med 2015 Apr;48(4):452-455. [doi: 10.1016/j.amepre.2014.10.010] [Medline: 25576494]
65. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, NJ, US: Prentice-Hall, Inc; 1986.
66. Luszczynska A, Schwarzer R. Changing behavior using social cognitive theory. In: Hagger MS, Cameron LD, Hamilton K, Hankonen N, Lintunen T, editors. The Handbook of Behavior Change, 1st Ed. Cambridge, MA: Cambridge University Press; 2020:32-45.
67. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. Am J Prev Med 2007 May;32(5 Suppl):112-118 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.022] [Medline: 17466815]

68. Rogers LQ, Courneya KS, Oster RA, Anton PM, Robbs RS, Forero A, et al. Physical activity and sleep quality in breast cancer survivors: a randomized trial. Med Sci Sports Exerc 2017 Oct;49(10):2009-2015 [FREE Full text] [doi: 10.1249/MSS.0000000000001327] [Medline: 28538261]
69. American Medical Association. 2020. URL: https://www.ama-assn.org/ [accessed 2020-11-20]
70. Gottesmann C, Gottesman I. The neurobiological characteristics of rapid eye movement (REM) sleep are candidate endophenotypes of depression, schizophrenia, mental retardation and dementia. Prog Neurobiol 2007 Mar;81(4):237-250. [doi: 10.1016/j.pneurobio.2007.01.004] [Medline: 17350744]
71. Lavie P, Pillar G, Malhotra A. Sleep Disorders Handbook: A Handbook for Clinicians. Boca Raton, Florida, United States: CRC Press; 2002.
72. Dao KP, De Cocker K, Tong HL, Kocaballi AB, Chow C, Laranjo L. Smartphone-delivered ecological momentary interventions based on ecological momentary assessments to promote health behaviors: systematic review and adapted checklist for reporting ecological momentary assessment and intervention studies. JMIR Mhealth Uhealth 2021 Nov 19;9(11):e22890 [FREE Full text] [doi: 10.2196/22890] [Medline: 34806995]
73. Carney CE, Buysse DJ, Ancoli-Israel S, Edinger JD, Krystal AD, Lichstein KL, et al. The consensus sleep diary: standardizing prospective sleep self-monitoring. Sleep 2012 Feb 01;35(2):287-302 [FREE Full text] [doi: 10.5665/sleep.1642] [Medline: 22294820]
74. Howick J, Chalmers I, Glasziou P, Greenhalgh T, Heneghan C, Liberati A, et al. The 2011 Oxford CEBM levels of evidence (Introductory document). Oxford Centre for Evidence-Based Medicine. URL: https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebm-levels-of-evidence [accessed 2020-09-26]
75. Buysse DJ, Ancoli-Israel S, Edinger JD, Lichstein KL, Morin CM. Recommendations for a standard research assessment of insomnia. Sleep 2006 Sep;29(9):1155-1173. [doi: 10.1093/sleep/29.9.1155] [Medline: 17040003]
76. James E, Freund M, Booth A, Duncan MJ, Johnson N, Short CE, et al. Comparative efficacy of simultaneous versus sequential multiple health behavior change interventions among adults: a systematic review of randomised trials. Prev Med 2016 Dec;89:211-223. [doi: 10.1016/j.ypmed.2016.06.012] [Medline: 27311332]
77. Lippke S. Modelling and supporting complex behavior change related to obesity and diabetes prevention and management with the compensatory carry-over action model. J Diabetes Obes 2014;1(2):1-5 [FREE Full text] [doi: 10.15436/2376-0494.14.009]
78. Kline CE. The bidirectional relationship between exercise and sleep: implications for exercise adherence and sleep improvement. Am J Lifestyle Med 2014;8(6):375-379 [FREE Full text] [doi: 10.1177/1559827614544437] [Medline: 25729341]
79. Rayward AT, Burton NW, Brown WJ, Holliday EG, Plotnikoff RC, Duncan MJ. Associations between changes in activity and sleep quality and duration over two years. Med Sci Sports Exerc 2018 Dec;50(12):2425-2432. [doi: 10.1249/MSS.0000000000001715] [Medline: 30001222]
80. Persky I, Spring B, Vander Wal JS, Pagoto S, Hedeker D. Adherence across behavioral domains in treatment promoting smoking cessation plus weight control. Health Psychol 2005 Mar;24(2):153-160 [FREE Full text] [doi: 10.1037/0278-6133.24.2.153] [Medline: 15755229]
81. Prochaska JJ, Sallis JF. A randomized controlled trial of single versus multiple health behavior change: promoting physical activity and nutrition among adolescents. Health Psychol 2004 May;23(3):314-318. [doi: 10.1037/0278-6133.23.3.314] [Medline: 15099173]
82. Taylor WC, Hepworth JT, Lees E, Cassells A, Gousse Y, Sweeney MM, et al. Readiness to change physical activity and dietary practices and willingness to consult healthcare providers. Health Res Policy Syst 2004 Jun 10;2(1):2 [FREE Full text] [doi: 10.1186/1478-4505-2-2] [Medline: 15191611]
83. Fleig L, Kerschreiter R, Schwarzer R, Pomp S, Lippke S. 'Sticking to a healthy diet is easier for me when I exercise regularly': cognitive transfer between physical exercise and healthy nutrition. Psychol Health 2014 Jul;29(12):1361-1372. [doi: 10.1080/08870446.2014.930146] [Medline: 24894668]
84. Baumeister RF, Bratslavsky E, Muraven M, Tice DM. Ego depletion: is the active self a limited resource? J Personal Soc Psychol 1998;74(5):1252-1265. [doi: 10.1037/0022-3514.74.5.1252]
85. Baumeister RF, Muraven M, Tice DM. Ego depletion: a resource model of volition, self-regulation, and controlled processing. Soc Cognit 2000 Jun;18(2):130-150. [doi: 10.1521/soco.2000.18.2.130]
86. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on internet-based mental health interventions: a systematic review. Internet Intervent 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
87. Michie S, Thomas J, Johnston M, Aonghusa PM, Shawe-Taylor J, Kelly MP, et al. The Human Behaviour-Change Project: harnessing the power of artificial intelligence and machine learning for evidence synthesis and interpretation. Implement Sci 2017 Oct 18;12(1):121 [FREE Full text] [doi: 10.1186/s13012-017-0641-5] [Medline: 29047393]
88. Irish LA, Kline CE, Gunn HE, Buysse DJ, Hall MH. The role of sleep hygiene in promoting public health: a review of empirical evidence. Sleep Med Rev 2015 Aug;22:23-36 [FREE Full text] [doi: 10.1016/j.smrv.2014.10.001] [Medline: 25454674]
89. American Academy of Sleep Medicine. International Classification of Sleep Disorders: Diagnostic & Coding Manual. Darien, Illinois, United States: American Academy of Sleep Medicine; 2005.

90. Morin C. Insomnia: Psychological Assessment and Management. New York: Guilford Press; 1993.
91. Murawski B, Plotnikoff RC, Lubans DR, Rayward AT, Brown WJ, Vandelanotte C, et al. Examining mediators of intervention efficacy in a randomised controlled m-health trial to improve physical activity and sleep health in adults. Psychol Health 2020 Nov 27;35(11):1346-1367. [doi: 10.1080/08870446.2020.1756288] [Medline: 32456468]

Abbreviations

BCT: behavior change technique
BCTTv1: BCT Taxonomy version 1
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
RQ: research question
SD: standard deviation

Edited by L Buis; submitted 10.09.21; peer-reviewed by Z Liang, S Rostam Niakan Kalhori; comments to author 11.10.21; revised version received 21.12.21; accepted 07.01.22; published 04.04.22.

Please cite as:
Arroyo AC, Zawadzki MJ
The Implementation of Behavior Change Techniques in mHealth Apps for Sleep: Systematic Review
JMIR Mhealth Uhealth 2022;10(4):e33527
URL: https://mhealth.jmir.org/2022/4/e33527
doi:10.2196/33527
PMID:35377327

Review

The Effectiveness of Combining Nonmobile Interventions With the Use of Smartphone Apps With Various Features for Weight Loss: Systematic Review and Meta-analysis

Jumana Antoun[1], MD, PhD; Hala Itani[1], MD; Natally Alarab[1], MD; Amir Elsehmawy[1], MD

American University of Beirut, Beirut, Lebanon

Corresponding Author:
Jumana Antoun, MD, PhD
American University of Beirut
Riad El-Solh
Beirut, 110236
Lebanon
Phone: 961 3486509
Email: ja46@aub.edu.lb

Abstract

Background: The effectiveness of smartphone apps for weight loss is limited by the diversity of interventions that accompany such apps. This research extends the scope of previous systematic reviews by including 2 subgroup analyses based on nonmobile interventions that accompanied smartphone use and human-based versus passive behavioral interventions.

Objective: The primary objective of this study is to systematically review and perform a meta-analysis of studies that evaluated the effectiveness of smartphone apps on weight loss in the context of other interventions combined with app use. The secondary objective is to measure the impact of different mobile app features on weight loss and mobile app adherence.

Methods: We conducted a systematic review and meta-analysis of relevant studies after an extensive search of the PubMed, MEDLINE, and EBSCO databases from inception to January 31, 2022. Gray literature, such as abstracts and conference proceedings, was included. Working independently, 2 investigators extracted the data from the articles, resolving disagreements by consensus. All randomized controlled trials that used smartphone apps in at least 1 arm for weight loss were included. The weight loss outcome was the change in weight from baseline to the 3- and 6-month periods for each arm. Net change estimates were pooled across the studies using random-effects models to compare the intervention group with the control group. The risk of bias was assessed independently by 2 authors using the Cochrane Collaboration tool for assessing the risk of bias in randomized trials.

Results: Overall, 34 studies were included that evaluated the use of a smartphone app in at least 1 arm. Compared with controls, the use of a smartphone app–based intervention showed a significant weight loss of –1.99 kg (95% CI –2.19 to –1.79 kg; I^2=81%) at 3 months and –2.80 kg (95% CI –3.03 to –2.56 kg; I^2=91%) at 6 months. In the subgroup analysis, based on the various intervention components that were added to the mobile app, the combination of the mobile app, tracker, and behavioral interventions showed a statistically significant weight loss of –2.09 kg (95% CI –2.32 to –1.86 kg; I^2=91%) and –3.77 kg (95% CI –4.05 to –3.49 kg; I^2=90%) at 3 and 6 months, respectively. When a behavioral intervention was present, only the combination of the mobile app with intensive behavior coaching or feedback by a human coach showed a statistically significant weight loss of –2.03 kg (95% CI –2.80 to –1.26 kg; I^2=83%) and –2.63 kg (95% CI –2.97 to –2.29 kg; I^2=91%) at 3 and 6 months, respectively. Neither the type nor the number of mobile app features was associated with weight loss.

Conclusions: Smartphone apps have a role in weight loss management. Nevertheless, the human-based behavioral component remained key to higher weight loss results.

(JMIR Mhealth Uhealth 2022;10(4):e35479) doi:10.2196/35479

KEYWORDS

obesity; weight loss; mobile app; self-monitoring; behavioral; tracker; behavioral coaching; coach; dietitian; mobile phone

Introduction

Background

Obesity has become a major, rising health epidemic worldwide. As a complex multifactorial disease, the management of obesity is challenging because there is no single effective treatment. Of late, there has been great interest in using apps for weight loss. Mobile apps were effective for weight loss [1-6] by using different behavior change techniques to a certain extent [1-6]. These behavior change techniques include intention formation, goal setting, barrier identification, problem solving, planning, general encouragement, self-monitoring of behavior, feedback on performance, social support, and social comparison [4,7].

On the basis of recent research, mobile apps help users to adhere to self-monitoring and weight loss goals better than the traditional pen-and-paper methods and other mobile health interventions (web-based or PDA) [8-12]. In 2015, Mateo et al [13] conducted the first meta-analysis that focused on mobile apps and found a modest weight loss of –1.04 kg (95% CI –1.75 to –0.34 kg; I^2=41%) among mobile app users. In 2014 and 2015, similar results were found by Khokhar et al [14], Hutchesson et al [15], and Liu et al [16] after they expanded the inclusion criteria to include email, SMS text messaging, monitoring devices, and smartphones. Cai et al [17] observed similar findings when they measured the effect of mobile apps in patients with diabetes mellitus. In 2020, Islam et al [18] updated the literature review and extended the scope of the previous meta-analysis performed by Mateo et al [13] by including more subgroup analyses.

Most of the interventions are smartphone apps combined with other behavioral nonmobile interventions; yet, it's unclear whether the app's effect on weight loss is due solely to its use or to the addition of the behavioral component. If the behavioral component relies on human coaches and personalized feedback by dietitians, this will affect the scalability of the mobile app used for weight loss. Personalized feedback provided by mobile apps has proven to be an essential feature of such apps because the feedback increases users' logging-in frequency and engagement with the apps [19-23]. Personalized feedback from an interventionist or professional also affected the results positively [24-26]. Thus, the combination of mobile app use with in-person contacts such as coaching or counseling sessions, interventionist feedback, web-based chatting, or telephone calls with professionals was more effective than mobile app use alone [19,24,27-29].

Objectives

The aim of this meta-analysis is to evaluate the effectiveness of mobile app interventions alone or in combination with other behavioral interventions on weight loss. Although Lyzwinski [30] analyzed the intervention components of mobile devices in a narrative review, the author did not compare the effect of the various components on weight loss. This research extends the scope of the previous systematic reviews by including 2 subgroup analyses based on nonmobile interventions that accompanied smartphone use and human-based versus passive behavioral interventions. The results are organized according to duration because it is inaccurate to compare weight loss at 3 months with weight loss at 6 or 12 months. The secondary outcomes of the study include the impact of different mobile app features on weight loss and mobile app adherence.

Methods

This systematic review of the literature and quantitative meta-analysis was conducted following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [31] to measure the effectiveness of mobile app interventions alone or in combination with other behavioral interventions on weight loss. The secondary outcomes of the study include the impact of different mobile app features on weight loss and adherence.

Protocol and Registration

The protocol was not registered.

Search Strategy

The PubMed, MEDLINE, and EBSCO databases were searched for relevant studies published between the database inception date and January 31, 2022. The search strategy incorporated keywords. The terms used included *weight loss*, *obesity*, *overweight*, *smartphone*, *mobile phone*, *cell phone*, *mHealth*, *eHealth*, and *adherence*. The search was then filtered to studies involving randomized controlled trials. All previous systematic reviews and meta-analyses were researched to find further missing studies. EndNote X9 (Clarivate) was used to remove duplicate publications and for screening purposes (for further information on the search strategy, please see the example provided in Multimedia Appendix 1). Multimedia Appendix 2 includes a summary of the reasons for excluding articles.

Study Selection

To investigate the effectiveness of nonapp interventions combined with smartphone apps, studies were eligible if (1) the design included randomized controlled trials, (2) they included the use of a smartphone app in at least 1 arm, (3) weight loss was an outcome, and (4) the population consisted of adults. There was no restriction on the population regarding overweight versus obesity or being diagnosed with chronic diseases, the language or year of publication, length of interventions, or follow-up duration. Gray literature, such as abstracts and conference proceedings, was included. We also searched the lists of references of the articles that we included.

On the basis of the eligibility criteria, 2 research team members (JA and HI) independently screened all the articles by study title and abstract. If the information listed in the title or abstract was insufficient to determine the study's relevance, the full text of the study was selected to be reassessed later. Next, each member further screened the selected studies at the full-text level. Any disagreements were resolved by consensus.

Data Collection Process and Data Items

Working independently, 2 investigators (JA and HI) extracted the data from the articles, resolving disagreements by consensus. A form developed by using the KoBo toolbox (Kobo Inc) was used to extract data from eligible research papers, including

digital object identifier, the title of the study, year of publication, type of article, country of study, population, sample size, trial name, trial period, number of arms, and details of each arm. Mean body weight changes were recorded from baseline to the end of the trial with SDs and adherence-related outcomes. SD or 95% CI was recorded if available. Neither authorship nor publication journal nor study results were blinded for data extraction.

Risk of Bias in Individual Studies

The risk of bias was assessed independently by 2 authors (NA and AE) using the Cochrane Collaboration tool for assessing the risk of bias in randomized trials [31]. The tool covers the following bias domains: selection bias, performance bias, detection bias, attrition bias, and reporting bias. Each author independently judged each domain as having a low, unclear, or high risk of bias. When differences of opinion arose between the 2 evaluators, the item was discussed until a consensus was reached. JA randomly selected a few articles to assess the risk of bias and compared the results with those of the 2 primary evaluators to consolidate the assessments further.

Measured Outcomes

The study's primary outcome is the mean weight change (measured in kilograms) from baseline to 3, 6, and 12 months. All outcomes were recorded if the study measured the outcome at multiple points. Adherence measures were examined, and a systematic review of the literature was performed.

Data Analysis and Synthesis of Results

The meta-analysis evaluated whether smartphone app interventions were effective on their own or whether other behavior interventions were necessary for weight loss. A fixed effect model was used to obtain the overall effect size across included studies and its associated 95% CI. On the basis of the studies chosen, the outcome of weight loss was measured as the weight change from baseline to the 3-, 6-, or 12-month periods. When SD was not mentioned, the variance was calculated from the 95% CI. We examined heterogeneity using the I^2 test, which describes the percentage of variability in effect size estimates because of heterogeneity rather than sampling error. The statistical analyses were performed using Review Manager software (version 5.4; Cochrane Training).

In studies with more than one arm that included an app, inverse variance meta-analysis was used to produce an overall effect size across all treatment arms, creating a single intervention-versus-control comparison for each study. The exact process was performed when there was more than one control arm.

All the interventions used were reviewed by 2 authors (JA and HI), who then grouped them into the following categories: smartphone apps, trackers (weighing scale, step tracker, or bite counter), behavioral therapy or advice (podcasts, telephone calls, booklets, SMS text messages, and in-person meetings), feedback (SMS text messages, email, oral, or written by a coach), self-monitoring, social support (social media or web-based forum), meal replacement, and financial incentives. Human-based active behavioral coaching or feedback included in-person meetings, interaction with the interventionist through Twitter or chat feature of the app, and tailored feedback from a coach or interventionist using telephone calls, SMS text messages, or group sessions. The passive behavioral component included passive standardized behavioral messages as part of the app, Facebook, or podcasts. The apps were also classified according to the following features: self-monitoring, education, feedback, social support, rewards, and gamification. On the basis of these features, 2 associations were later analyzed. Using an independent 2-tailed *t* test, the first analysis examined the association between weight loss and each app feature. The second analysis, using 1-way analysis of variance, studied the association between weight loss and the number of features. We aim to measure the association between adherence percentage and app features. Nevertheless, the definition of adherence was not homogeneous, and we ultimately conducted a systematic review of the adherence outcome.

Concerning the intervention period, it included both active treatment and follow-up. If provided, the baseline weight of the intervention group was noted; however, if it was not available, the average weight of participants in the whole cohort was captured from the demographics table.

Results

Study Selection and Characteristics

The search strategy enabled us to compile 1081 articles from different resources, of which 34 (3.14%) were included in this meta-analysis (Figure 1). The studies selected were published between 2011 and January 31, 2022; however, 68% (23/34) of the articles were published in the last 5 years (2017-2021). The studies were conducted in the United States (22/34, 65%), Australia and New Zealand (3/34, 9%), Europe (4/34, 12%; Germany: 1/4, 25%; Spain: 1/4, 25%; and the United Kingdom: 2/4, 50%), and Asia (5/34, 15%; Japan: 1/5, 20%; Singapore 1/5, 20%; and South Korea: 3/5, 60%). The sample size ranged from 16 to 440, with a mean of 113.09 (SD 94.1). The population in the studies ranged from men and women from the general population to adults at risk for diabetes as well as adults with diseases such as cardiology issues, diabetes, or metabolic syndrome. Of the 34 studies, 7 (20%) used a theoretical framework: social cognitive theory (3/7, 43%), social cognitive theory and transtheoretical model (2/7, 29%), and social cognitive theory and self-efficacy theory (2/7, 29%). Tables 1 and 2 present a summary of the characteristics of the included studies.

Figure 1. Flowchart of selection of the articles based on PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. RCT: randomized controlled trial.

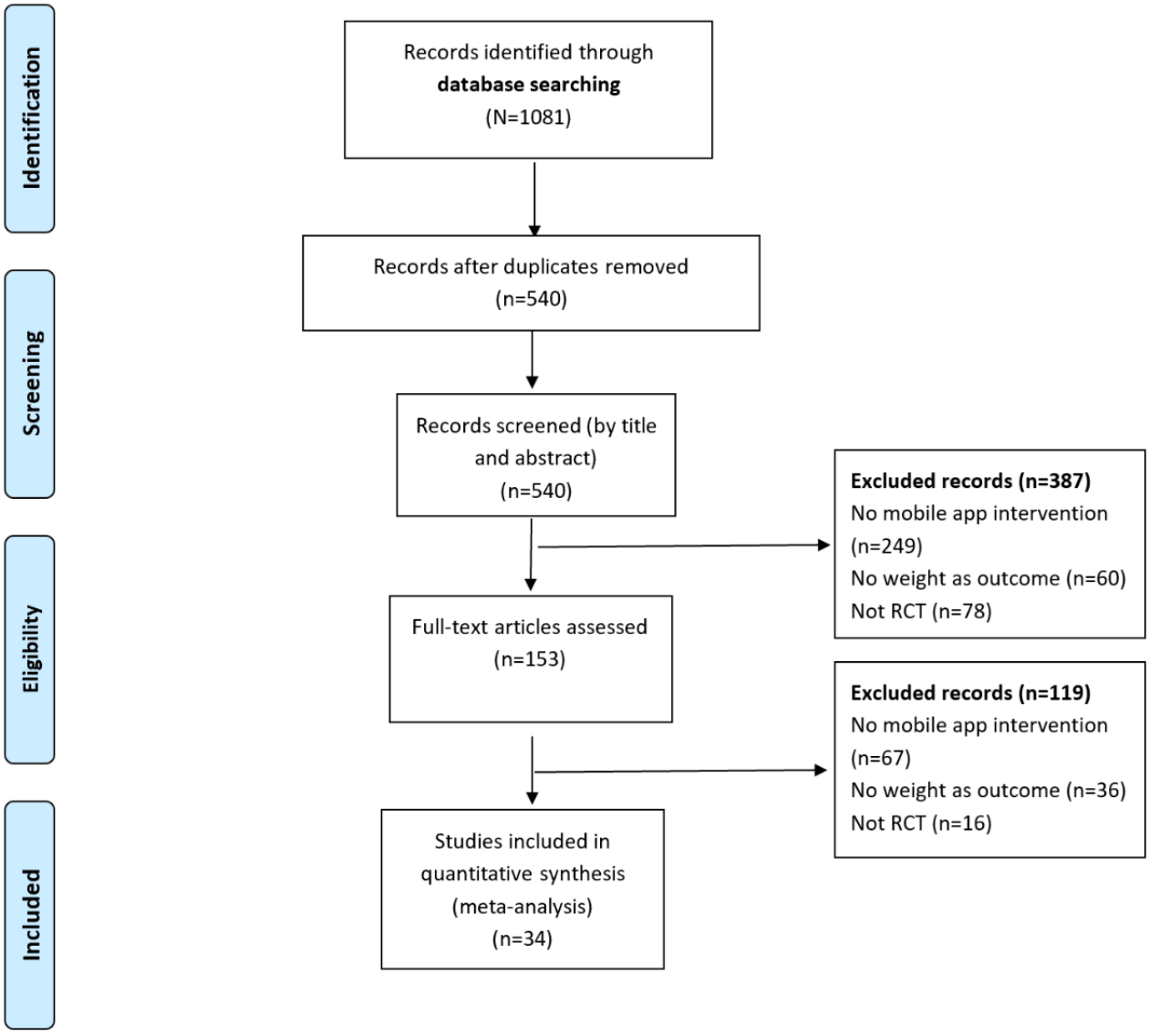


Table 1. Description of the population, sample size, and baseline demographics of the included studies (N=34).

Study	Country	Population of study	Sample size, n	Trial period (weeks)	Age (years), mean (SD) or mean (95% CI)	Baseline BMI (kg/m^2), mean (SD) or mean (95% CI)	Baseline weight (kg). mean (SD) or mean (95% CI)
Bender and Cooper, 2017 [32]	United States	Adults with diabetes	45	24	57.6 (9.8)	30.1 (4.6)	75.8 (15.4)
Fukuoka et al, 2015 [33]	United States	Adults with prediabetes	61	20	55.2 (9.0)	33.3 (6.0)	86.2 (18.5)
Whitelock et al, 2019 [34]	United Kingdom	Adults	107	8	42.8 (10.5)	35.9 (6.8)	100.5 (20.4)
Vaz et al, 2018 [35]	United States	Adults	28	24	39.5 (3.71)	34.5 (1.3)	94.3 (3.42)
Thompson-Felty and Johnston, 2017 [26]	United States	Adults	30	8	__[a]	—	—
Rogers et al, 2016 [36]	United States	Adults	39	24	39.9 (11.5)	39.5 (2.8)	111.5 (11.5)
Svetkey et al, 2015 [29]	United States	Adults	365	96	29.4 (4.3)	35.2 (7.8)	101 (23.7)
Thomas et al, 2019 [37]	United States	Adults	276	72	55.1 (9.9)	35.2 (5)	95.9 (17.0)
Brindal et al, 2013 [38]	Australia	Adult women	58	8	42 (—)	34 (—)	92.4 (14.7)
Laing et al, 2014 [39]	United States	Adult primary care patients	212	24	43.3 (14.3)	33.4 (7.09)	—
Spring et al, 2017 [28]	United States	Adults	96	48	39.3 (11.7)	34.6 (3.0)	94.8 (12.4)
Shin et al, 2017 [40]	South Korea	Adult men	105	12	27.8 (—)	29.8 (—)	91.4 (10.6)
Ross and Wing, 2016 [27]	United States	Adults	80	24	51.1 (11.7)	33.0 (3.4)	—
Gilmore et al, 2017 [41]	United States	Postpartum women	40	16	26.0 (5.2)	31.3 (3.2)	83.8 (13.5)
Tanaka et al, 2018 [42]	Japan	Adults	112	12	45.6 (10.2)	28.0 (3.3)	83.1 (11.1)
Allen et al, 2013 [43]	United States	Adults	68	24	44.9 (11.1)	34.3 (3.9)	97.3 (16.2)
Stephens et al, 2017 [44]	United States	Adults	62	12	20.0 (18.0-25.0)	28.5 (25.0-40.4)	82.8 (61-117.5)
Hales et al, 2016 [45]	United States	Adults	51	12	46.2 (12.4)	34.7 (6.0)	102.1 (91.9-112.2)
Hartman et al, 2016 [46]	United States	Adult women with elevated risk for breast cancer	54	24	59.5 (5.6)	31.9 (3.5)	86.3 (10.2)
Haufe et al, 2019 [47]	Germany	Adults with metabolic syndrome	314	24	48.1 (8.1)	33.3 (5.4)	106.7 (19.1)
Turner-McGrievy et al, 2017 [48]	United States	Adults who were overweight	81	24	48.6 (11.7)	33.4 (4.8)	—
Jospe et al, 2017 [49]	New Zealand	Adults with overweight or obesity	250	48	44.4 (10.2)	33.5 (4.5)	99.1 (17.3)
Burke et al, 2017 [50]	United States	Adults	39	12	44.85 (12.75)	33.76 (4.28)	93.15 (15.89)
Lee et al, 2019 [51]	South Korea	Adults with metabolic syndrome	129	24	30.59 (—)	—	71.6 (12.2)
Turner-McGrievy and Tate, 2011 [52]	United States	Adults	96	24	42.6 (10.7)	32.9 (4.8)	—
Monroe et al, 2019 [53]	United States	Adults	36	48	44.67 (8.96)	36.22 (7.53)	97.78 (21.04)
Choi et al, 2019 [54]	United States	Adult patients with cardiology issues	100	24	57.2 (1.8)	29.5 (0.6)	84.8
Evangelista et al, 2018 [55]	United States	Adults with heart failure	16	12	52.3 (8.5)	—	—

Study	Country	Population of study	Sample size, n	Trial period (weeks)	Age (years), mean (SD) or mean (95% CI)	Baseline BMI (kg/m^2), mean (SD) or mean (95% CI)	Baseline weight (kg). mean (SD) or mean (95% CI)
Kurtzman et al, 2018 [56]	United States	Adults	196	36	42.3 (11.5)	36.0 (5.2)	102.0 (18.8)
Carter et al, 2013 [12]	United Kingdom	Adults	128	24	41.2 (8.5)	33.7 (4.2)	96.4 (16)
Duncan et al, 2020 [57]	Australia	Adults	116	48	44.5 (10.4)	31.7 (3.9)	90.7 (14.3)
Lim et al, 2021 [58]	Singapore	Adults with diabetes	204	24	51.2 (9.7)	84.0 (12.6)	30.6 (4.3)
Ahn et al, 2020 [59]	South Korea	Adults	50	6	26.0 (4.8)	77.1 (11.5)	26.7 (2.7)
Lugones-Sanchez et al, 2020 [60]	Spain	Adults	440	12	47.4 (10.0)	89.7 (13.1)	32.8 (3.3)

[a]Not available.

Table 2. Description of the study arms, app features, and theoretical frameworks (N=34).

Study	Number of arms	Intervention description	Control description	App features	Commercial app	Theoretical framework of the app
Bender and Cooper, 2017 [32]	2	App+tracker+social support (Facebook)+behavior or advice (in-person meetings)	Tracker+waitlist	Self-monitoring	__[a]	Social cognitive theory and transtheoretical model for health behavior change
Fukuoka et al, 2015 [33]	2	App+tracker+behavior or advice (in-person meetings)	Tracker+behavior or advice (booklet)	Self-monitoring+education+feedback+gamification	—	—
Whitelock et al, 2019 [34]	2	App+behavior or advice (booklet+SMS text messages)	Behavior or advice (booklet+SMS text messages)	Self-monitoring+education+feedback+rewards	—	—
Vaz et al, 2018 [35]	2	App+tracker	Behavior or advice (in-person meetings)	Self-monitoring+education+social support+rewards	—	—
Thompson-Felty and Johnston, 2017 [26]	3	All 3 arms included an app: arm 1: app; arm 2: app+feedback; arm 3: app	No control	Arm 1track pictures of foods; Arm 2: track picture of foods+feedback; arm3: track pictures of foods+education	—	—
Rogers et al, 2016 [36]	3	App+tracker+behavior or advice (telephone call+booklet)	One arm: tracker+behavior or advice (telephone call+booklet)+web-based self-monitoring; second arm: behavior or advice (in-person meetings+booklet)+feedback (oral by coach)+paper-based self-monitoring	Self-monitoring	—	—
Svetkey et al, 2015 [29]	3	Arm 1: app+tracker; arm 2: app+tracker+social support (social buddy)+behavior or advice (in-person meetings)	Behavior or advice (booklet)	Arm 1: self-monitoring+feedback+social support+gamification; arm 2: self-monitoring	—	Social cognitive theory and transtheoretical model
Thomas et al, 2019 [37]	3	App+behavior or advice (in-person meetings)+feedback (oral and written by coach)	Arm 1: behavior or advice (in-person meetings)+feedback (oral and written by coach)+self-monitoring (paper diaries); arm 2: behavior or advice (dietary advice)+feedback (oral or written by coach)+self-monitoring (paper diaries)	Self-monitoring+education+social support	MyFitnessPal	—
Brindal et al, 2013 [38]	2	App+meal replacement	Meal replacement	Self-monitoring+education+feedback+rewards	—	—
Laing et al, 2014 [39]	2	App	Usual care	Self-monitoring+feedback+social support	MyFitnessPal	—
Spring et al, 2017 [28]	3	App+tracker+behavior or advice (in-person meetings)+feedback (telephone call by coach)	Arm 1: behavior or advice (in-person meetings)+feedback (telephone call by coach)+self-monitoring (paper diaries); arm 2: behavior or advice (DVDs)+self-monitoring (paper diaries)	Self-monitoring+social support	—	—

Study	Number of arms	Intervention description	Control description	App features	Commercial app	Theoretical framework of the app
Shin et al, 2017 [40]	3	Arm 1: app+tracker+behavior or advice (in-person meetings); arm 2: app+tracker+behavior or advice (in-person meeting)	Behavior or advice (in-person meetings)	Arm 1: self-monitoring+feedback+rewards; arm 2: self-monitoring	FitLife	—
Ross and Wing, 2016 [27]	3	Arm 1: app+tracker+behavior or advice (in-person meetings); arm 2: app+tracker+behavior or advice (in-person meetings and telephone calls)	Behavior or advice (in-person meetings)+tracker+self-monitoring (paper diaries)	Arms 1 and 2: self-monitoring+feedback	Fitbit	—
Gilmore et al, 2017 [41]	2	App+tracker	Usual care	Self-monitoring+education+feedback	—	—
Tanaka et al, 2018 [42]	2	App	Usual care	Self-monitoring+education+feedback+social support	FiNC	—
Allen et al, 2013 [43]	4	Arm 1: app; arm 2: app+behavior or advice (intensive in-person meetings); arm 3: app+behavior or advice (less intensive in-person meetings)	Behavior or advice (in-person meetings)	Arms 1, 2, and 3: self-monitoring+feedback+social support	Lose it!	Social cognitive theory
Stephens et al, 2017 [44]	2	App+behavior or advice (in-person meetings)+feedback (SMS text messages by health coach)	Behavior or advice (in-person session)	Self-monitoring+social support	Lose It!	Social cognitive theory and self-efficacy theory
Hales et al, 2016 [45]	2	Arm 1: app+behavior or advice (podcasts); arm 2: app+behavior or advice	No control	Arm 1: self-monitoring+feedback+social support+rewards; arm 2: self-monitoring	FatSecret	Social cognitive theory
Hartman et al, 2016 [46]	2	App+tracker+behavior or advice (telephone calls with coach)	Behavior or advice (telephone calls)	Self-monitoring	MyFitnessPal	—
Haufe et al, 2019 [47]	2	App+tracker+behavior or advice (in-person meetings)	Waitlist	Self-monitoring+education+feedback	—	—
Turner-McGrievy et al, 2017 [48]	2	App+tracker+behavior or advice (podcasts)	Tracker (bite counter)+behavior or advice (podcasts)	Self-monitoring	FatSecret	—
Jospe et al, 2017 [49]	5	App	Arm 1: behavior or advice (in-person meetings)+feedback (email); arm 2: behavior or advice (in-person meetings)+self-monitoring (hunger using capillary glucose monitor); arm 3: behavior or advice (monthly in-person meetings); arm 4: behavior or advice (in-person session at baseline)	Self-monitoring	MyFitnessPal	—
Burke et al, 2017 [50]	3	Arm 1: app; arm 2: app+(in-person meetings)+social support (Facebook); arm 3: app+social support (Facebook)	No control	Self-monitoring+feedback	Lose It!	—

Study	Number of arms	Intervention description	Control description	App features	Commercial app	Theoretical framework of the app
Lee et al, 2019 [51]	3	Arm 1: app; arm 2: app+behavior or advice (in-person meetings)	Usual care	Self-monitoring	—	—
Turner-McGrievy and Tate, 2011 [52]	2	App+behavior or advice (podcasts)+social support (Twitter)	Podcasts	Self-monitoring	FatSecret	Social cognitive theory
Monroe et al, 2019 [53]	2	Arm 1: app+tracker+behavior or advice (in-person meetings)+social support (support partners)+feedback (written by coach through website); arm 2: app+tracker+behavior or advice (in-person meetings)+feedback (written by coach through website)	No control	Self-monitoring+social support	MyFitnessPal	—
Choi et al, 2019 [54]	2	App+behavior or advice booklet (1-hour in-person meeting)	Behavior or advice (in-person meetings+telephone calls+booklets)	Self-monitoring+education+feedback+gamification	—	—
Evangelista et al, 2018 [55]	2	Arm 1: app+tracker+behavior or advice (in-person meetings)+feedback (SMS text messages by coach); arm 2: app+tracker+behavior or advice (in-person meetings)	No control	Self-monitoring	GetFit	—
Kurtzman et al, 2018 [56]	3	Arm 1: app+tracker+social support (support partners)+feedback (SMS text messages or email or both by primary care practitioner)+financial incentives; arm 2: app+tracker+social support (support partners)+financial incentives; arm 3: app+tracker+social support	No control	Self-monitoring	Withings Health Mate	—
Carter et al, 2013 [12]	3	App+social support (web-based forum)	Arm 1: social support (web-based forum)+self-monitoring (web-based); arm 2: social support (social forum)+self-monitoring (paper diaries)	Self-monitoring+feedback	My Meal Mate study mobile app	—
Duncan et al, 2020 [57]	3	Arm 1: app+tracker+advice or behavior (in-person counseling sessions); arm 2: same as arm 1+sleep goals	Waitlist	Self-monitoring+education+feedback	—	Social cognitive theory and self-efficacy theory
Lim et al, 2021 [58]	2	App+behavior or advice (dietitian contact through the app)	Not explained	Self-monitoring+education+feedback	—	—
Ahn et al, 2020 [59]	2	App	Paper diary	Self-monitoring	—	—
Lugones-Sanchez et al, 2020 [60]	2	App+tracker+behavior or advice (5-minute baseline session)	Standard counseling session (5-minute baseline session)	Self-monitoring	—	—

[a]Not available.

Assessment of the Risk

The studies selected for the meta-analysis used the following two types of analyses for the intervention results: intention-to-treat and per-protocol analyses. The intention-to-treat studies (n=21) were assessed for risk, and the per-protocol studies (n=13) were similarly evaluated. Among the 21 intention-to-treat studies, 28% (6/21) were determined to be low risk, 33% (7/21) had some concerns, and 38% (8/21) were determined to be high risk. The highest risk was related to outcome measurement because of the nature of the app and the consequent lack of blinding. Similarly, among the 13 per-protocol studies, 23% (3/13) were determined to be low risk, 38% (5/13) had some concerns, and 38% (5/13) were determined to be high risk. In the per-protocol studies, high risk was due to deviation from the intended interventions. Figure 2 [12,26-29,32-60] shows a summary of the risk of bias assessment of the included studies.

Figure 2. Summary of the risk of bias assessment of the included studies performed by using the Cochrane Collaboration tool [12,26-29,32-60].

Smartphone App Intervention and Weight Loss

Of the 34 included studies, 24 (71%) examined the effectiveness of smartphone apps on weight loss at 3 and 6 months, whereas 5 (15%) measured the outcomes at 12 months; of these 5, 2 (40%) did not include SD or 95% CI [49,56] and 1 (20%) included an app in all arms [53]. Consequently, the meta-analysis and subgroup analysis were performed for the 3- and 6-month outcomes. Compared with the control group, smartphone apps resulted in a pooled net estimate weight loss of –1.99 kg (95% CI –2.19 to –1.79 kg; I^2=81%) and –2.80 kg (95% CI –3.03 to –2.56 kg; I^2=90%) at 3 and 6 months, respectively (Figure 3 [27,28,32,33,36,39-42,44,48,52,58,60] and Figure 4 [12,27-29,35-37,39,43,46-48,51,52,54,57,58]). Subgroup analysis was performed based on the different interventions that accompanied the use of the mobile app. When compared with control, the combination of the mobile app, tracker, and behavioral interventions showed a statistically significant weight loss of –2.09 kg (95% CI –2.32 to –1.86 kg; I^2=87%) and –3.77 kg (95% CI –4.05 to –3.49 kg; I^2=90%) at 3 and 6 months, respectively (Figures 3 and 4). Another subgroup analysis was performed based on the type of behavioral interventions, human-based versus passive. When compared with control, only the combination of the mobile app with intensive behavior coaching or feedback by a human coach showed a statistically significant weight loss of –2.03 kg (95% CI –2.80 to –1.26 kg; I^2=83%) and –2.63 kg (95% CI –2.97 to –2.29 kg; I^2=91%) at 3 and 6 months, respectively (Figure 5 [27,28,32,33,36,42,44,48,50,52,55] and Figure 6 [27-29,35,36,43,46,47,51,52,54,57]). The funnel plots (Multimedia Appendix 3) were symmetrical, suggesting that there was no publication bias.

Figure 3. Forest plot of the effectiveness of mobile phone apps and the additional interventions on weight loss at 3 months [27,28,32,33,36,39-42,44,48,52,58,60]. IV: inverse variance method.

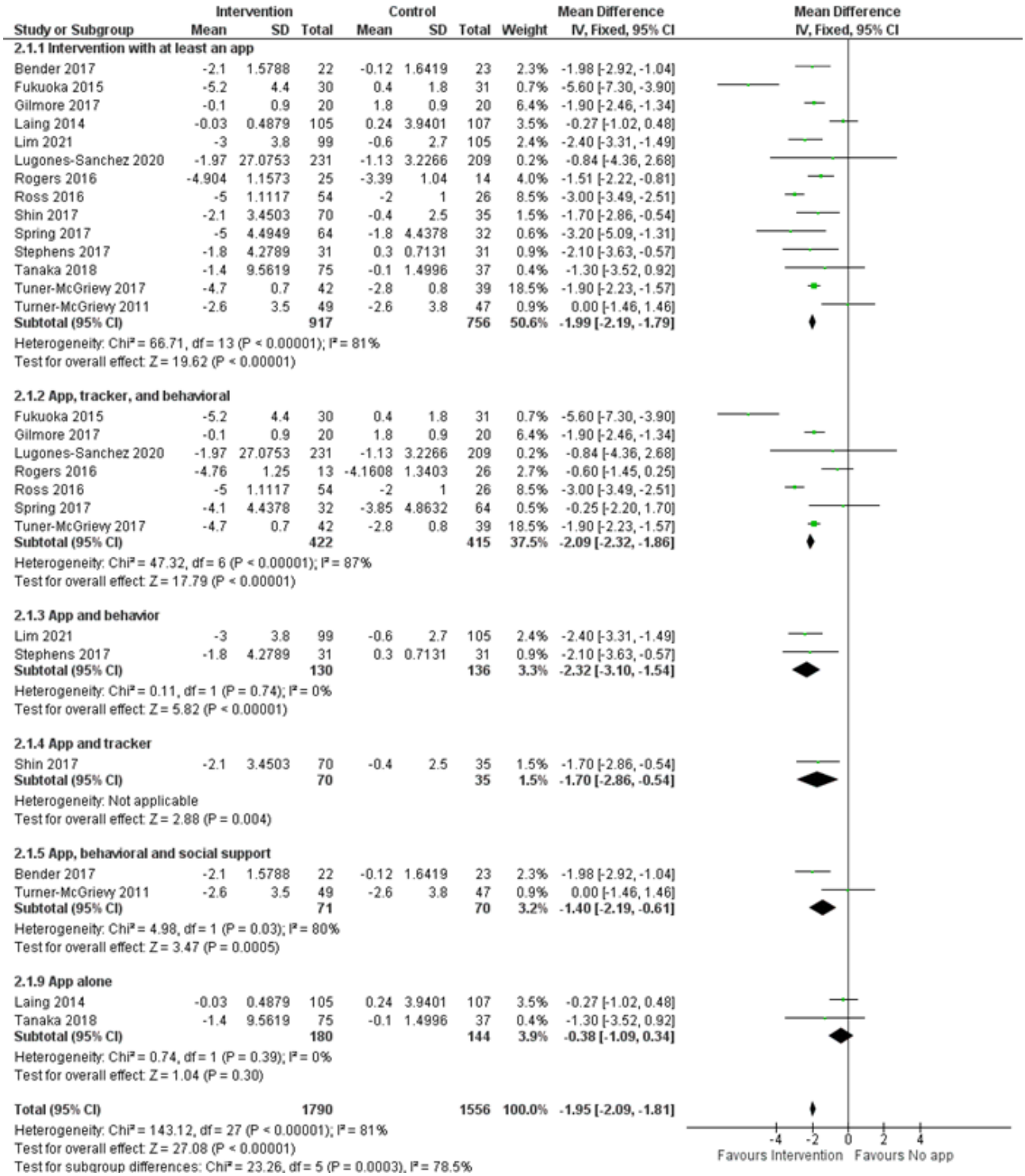


Figure 4. Forest plot of the effectiveness of mobile phone apps and the additional interventions on weight loss at 6 months [12,27-29,35-37,39,43,46-48,51,52,54,57,58]. IV: inverse variance method.

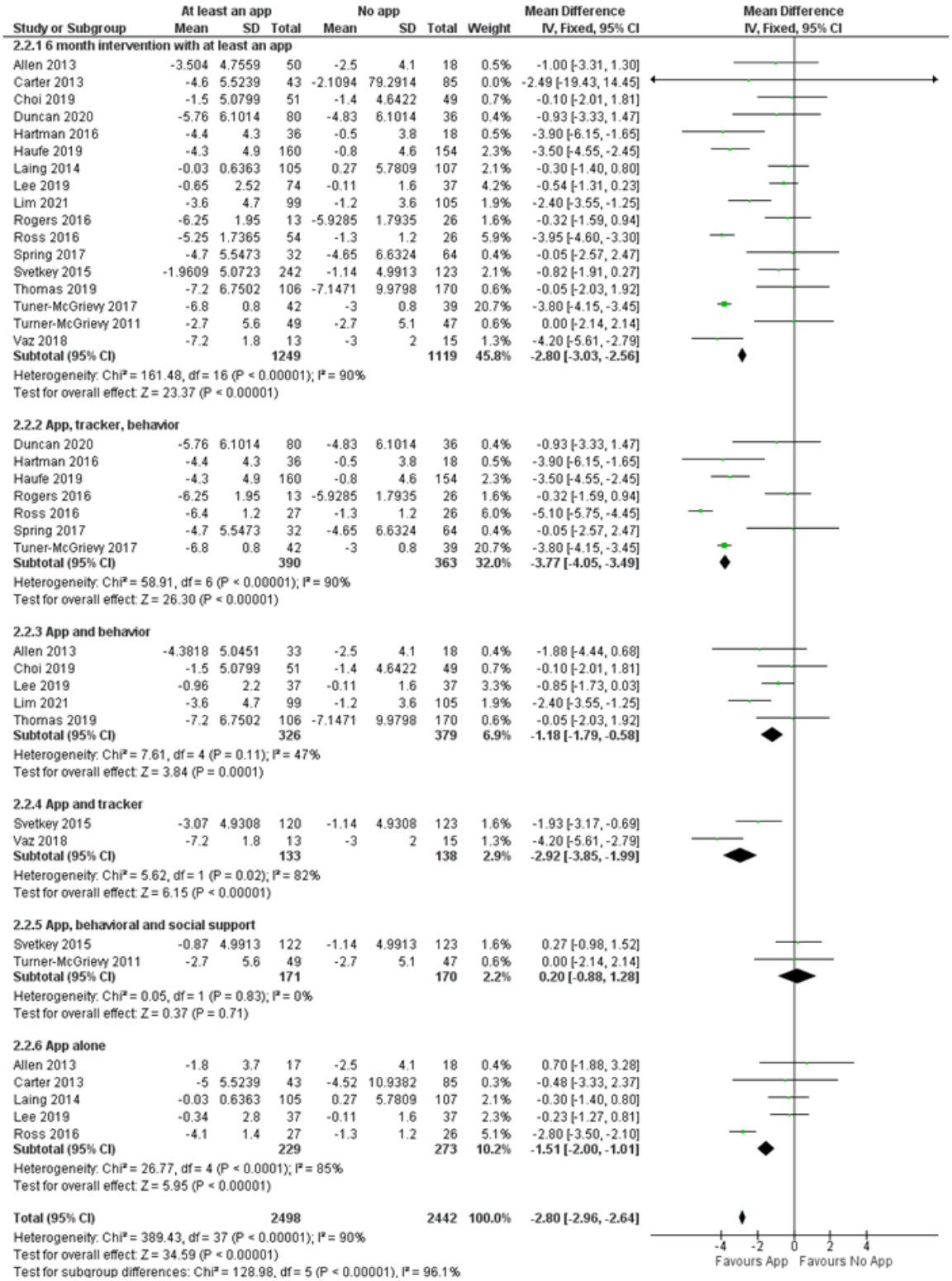


Figure 5. Subgroup analysis based on human-based versus passive behavioral interventions in combination with mobile app at 3 months [27,28,32,33,36,42,44,48,50,52,55]. IV: inverse variance method.

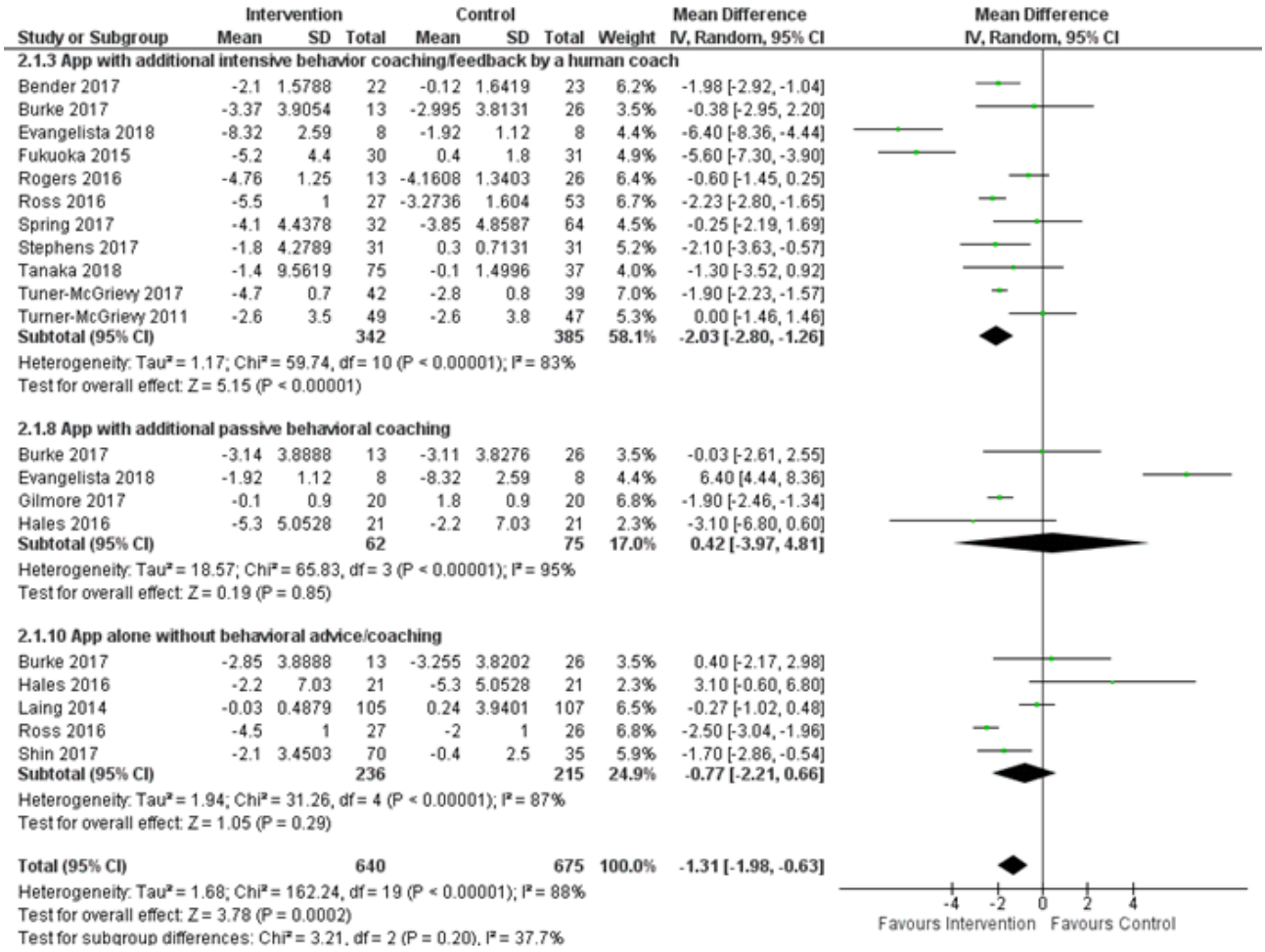
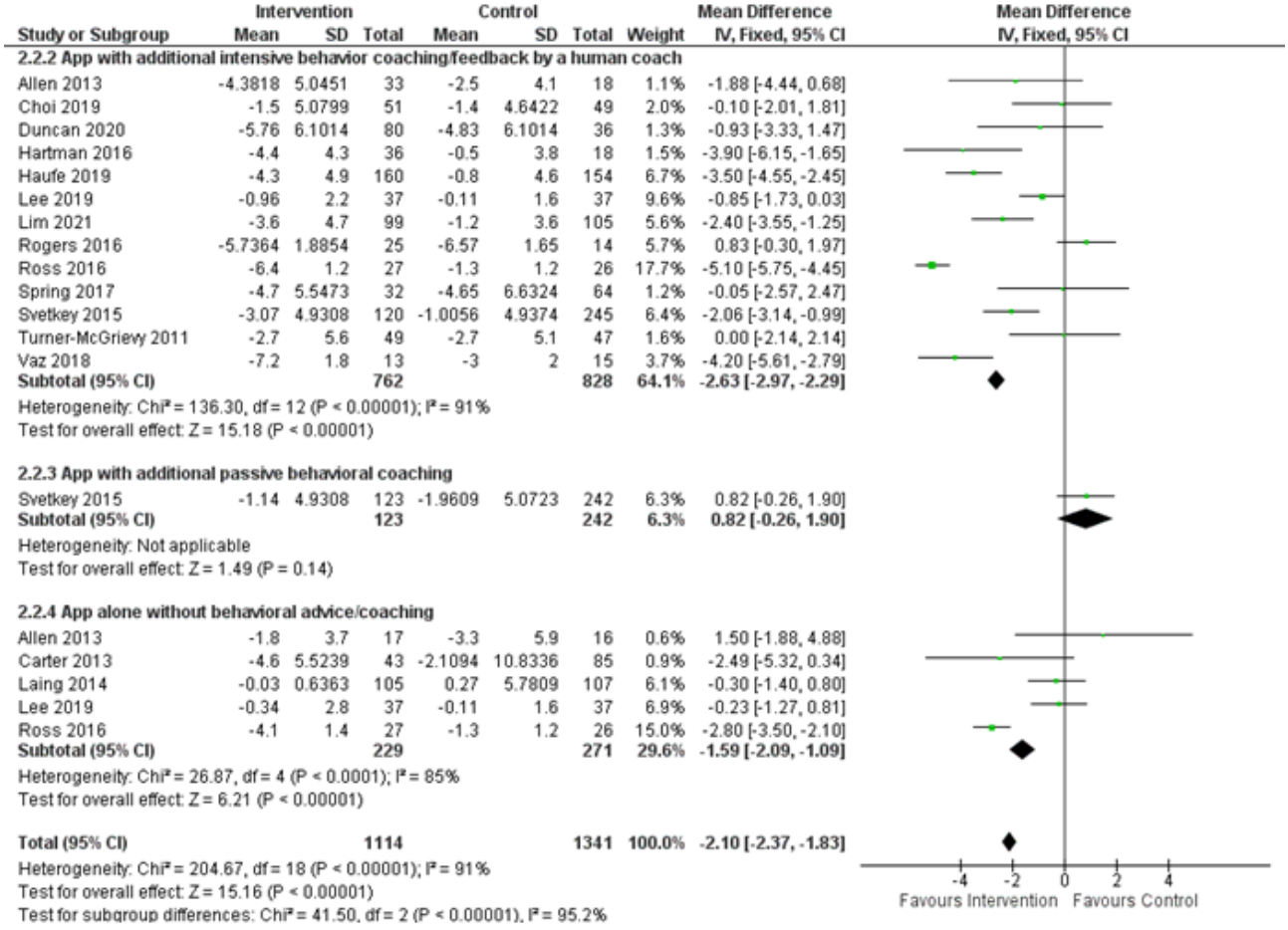


Figure 6. Subgroup analysis based on human-based versus passive behavioral interventions in combination with mobile app at 6 months [27-29,35,36,43,46,47,51,52,54,57]. IV: inverse variance method.



Characteristics of the Components of Both the Intervention and Control Arms

The studies included a total of 50 arms with an app and 35 control arms. Table 3 presents the characteristics of the various components of both the intervention and control arms. The app alone was used in 18% (9/50) of the intervention arms, whereas social and feedback interventions were not common, and financial incentives were the least used. The two most common combinations of interventions included the app, behavior component, and tracker (10/50, 20%) and app and behavior component (8/50, 16%). Among the 35 control arms, the behavior component was present in 21 (60%) and self-monitoring in 10 (29%), whereas no action was present in only 4 (11%). Multimedia Appendix 4 shows the details of the various components by study.

Table 3. Characteristics of the various components of both the intervention and control arms.

Components	Intervention arms (N=50), n (%)[a]	Control arms (N=35), n (%)[a]
App	50 (100)	N/A[b]
Behavior	32 (64)	21 (60)
Tracker	24 (50)	5 (14)
Social	10 (20)	2 (6)
Feedback	8 (16)	5 (14)
Financial incentives	2 (4)	0 (0)
Meal replacement	1 (2)	1 (3)
Self-monitoring	50 (100)	10 (29)
Usual care or waitlist	N/A	7 (20)

[a]Sum is more than 100% because arms could have more than 1 component.

[b]N/A: not applicable.

Features of the Mobile App and Weight Loss

Commercial mobile apps were used in 56% (28/50) of the study arms. Key features of the mobile apps included self-monitoring (50/50, 100%), feedback (20/50, 40%), education (15/50, 30%), social support (12/50, 24%), rewards (7/50, 14%), and gamification (3/50, 6%; Table 4). Two-thirds of the mobile apps included 1-2 features. There was no association between weight loss and any specific feature; neither was there an association between weight loss and the number of app features (data not shown).

Table 4. Frequency of app features (N=50).

	Frequency, n (%)
Feature	
Self-monitoring	50 (100)
Education	15 (30)
Feedback	20 (40)
Social support	12 (24)
Rewards	7 (14)
Gamification	3 (6)
Number of features per app	
1	22 (44)
2	9 (18)
3	12 (24)
4	7 (14)

Adherence to Smartphone Apps

In this meta-analysis, of the 34 included articles, 22 (65%) studied adherence to smartphone apps with diverse approaches to assessing adherence (Multimedia Appendix 5), leading to limitations in a direct comparison of their findings; thus, the results are described in a systematic review rather than a meta-analysis. Adherence data were extracted from each study, and a member of the research team (NA) coded the assessments based on the following four themes: (1) self-monitoring of weight, (2) self-monitoring of dietary intake, (3) self-monitoring of physical activity, and (4) interaction with the app. Each of these themes was defined differently in terms of measurement of adherence. Of the 22 articles that studied adherence to smartphone apps, 12 (55%) studies used more than one theme of adherence to weight loss apps; therefore, the total number of studies under each theme is not equal to the combined total number of studies included in this review.

Although dietary self-monitoring was the most commonly used adherence method among the studies (16/22, 73%), its measurement and analysis varied considerably. Studies defined adherence to dietary self-monitoring with a smartphone app either as recording any food or calorie intake [26,27,32,36,43,44,48,49,52,61] or a specific amount of calories [12,28,50] or a particular number of meals or entries [34,54] or both [37]. The frequency of dietary intake was based on the total number of days reported in percentages, ratios, or discrete numbers. Of the 22 studies, only 1 (5%) measured adherence as the percentage of participants logging food or calorie intake at least once per week [32]. The adherence rate to dietary intake ranged from 48% to 79% of the days when using self-monitoring.

App adherence for self-monitoring of weight and exercise was recorded in 36% (8/22) [27-29,32,37,38,49,54] and 27% (6/22) [27,28,36,44,52,54] of the studies, respectively. Common calculations included mean or percentage of daily recordings, recording at least once per week, or an average of days that participants recorded per week.

Of the 22 studies, 9 (41%) reported adherence as the frequency of the interaction with the app. Adherence was defined as wearing the wearable [32,33] or logging in [43,45] or opening the app, irrespective of the participant's use [39]. The frequency of wearing the wearable ranged from at least 4-5 days per week [32,33] to at least 8 hours per day [33]. In contrast, a group of studies considered counting the daily interaction with specific app components or features as a sign of adherence [29,34,38,61]. For example, some considered completing entries immediately after taking photographs of the meal a sign of adherence [34].

Some studies reported positive associations between adherence and weight change by using combined adherence measurements [28], tracking adherence with dietary intake app recording [12,26,27,48], and measuring adherence with physical activity app recording [44]. In some of the studies [29,33,34,50,52], the adherence percentage showed evident decline throughout the study period.

Discussion

Principal Findings

This meta-analysis aims to measure the effectiveness of smartphone app–based interventions on weight loss, considering the additional components available in the mobile app. Similar to previous meta-analyses [13,18], the use of mobile apps resulted in a small significant weight loss of –2.03 kg (95% CI

–2.57 to –1.5 kg; I^2=83%) at 3 months. Although the mobile apps included different behavioral strategies, all relied on self-monitoring and only one-third included more than two features. It is important to note that there was no association between weight loss and mobile app features. A review of mobile app features revealed that self-monitoring was most commonly used, whereas social support and personalized feedback were less commonly used [62]. Subgroup analysis integrating additional nonapp intervention components with the mobile app showed that use of the tracker and behavioral components resulted in the most significant weight loss of –3.77 kg (95% CI –4.05 to –3.49 kg) at 6 months. Human-based behavioral interventions were associated with weight loss of –2.63 kg (95% CI –2.97 to –2.29 kg) at 6 months.

The various app features were not associated with weight loss. The meta-analysis has shown significant heterogeneity among the different apps used and the additional nonapp intervention components, reaching 90% in some forest plots. It is difficult to determine the role of a mobile app in weight loss management beyond self-monitoring. Only a few studies (7/36, 19%) based their work on theoretical frameworks such as social cognitive theory, transtheoretical models for health behavior change, and self-efficacy theory. Both the transtheoretical framework and self-efficacy theories rely on the individual, supporting the importance of self-monitoring as the main feature of mobile apps. Only social cognitive theory addresses the importance of support; the study results highlighted the importance of support by human coaches. The behavioral components of the included studies in this subgroup analysis were mainly in-person meetings [28,32,33,40], with additional feedback or telephone calls by a coach [28,36]. There were numerous smartphone app features such as self-monitoring with additional feedback [32,36,48], gamification and awards [33,40], and social support [28]. In the form of an in-person meeting or telephone call, the human component, in combination with the user app interactions through self-monitoring and feedback, is crucial for weight loss. The need for the human component raises the question of whether artificial intelligence would raise the mobile apps to a new level in the management of weight loss [63], such as the use of chatbots [64,65], and whether users would accept such a mode of coaching [66].

Weight maintenance is defined as losing 5%-10% of body weight and maintaining this loss for at least 1 year. In this meta-analysis, the studies analyzed were mainly short term and lasted for a maximum of 1 year. Of the 36 studies included in this systematic review, 2 (6%) had a longer intervention duration, and neither showed any difference in weight from baseline between the app-based intervention and the control at 18 months [37] and 24 months [29]. In a systematic review, Varkevisser et al [67] have provided strong evidence that behavioral determinants such as self-monitoring of weight and eating predict weight loss maintenance. Moreover, web-based interventions were effective for weight maintenance [68]. It would be helpful to examine further how mobile apps can be a form of self-regulation and adherence that can help users maintain weight loss. As obesity is a long-term relapsing disease and mobile apps are cost-effective, further research should address whether mobile apps could play a role in weight loss maintenance.

Some studies in this review and the literature [69-72] suggest that greater adherence to self-monitoring has been associated with greater weight loss. However, many articles do not provide detailed measurements of adherence to self-monitoring in weight loss apps. Furthermore, the results across studies could not be accurately compared because of the numerous variations in measurement methods and the definitions of themes used to assess adherence. Irrespective of the measurement method used, adherence to self-monitoring decreased with time, emphasizing the importance of studying different app features and associated interventions that could have affected the participants' adherence. Of note, the variation in adherence measurements made it challenging to compare data across various studies. Hence, it is necessary to formulate a well-structured standard definition of adherence measures that can be used across future studies.

Strengths and Limitations

To our knowledge, this is the only meta-analysis that has performed a subgroup analysis based on the add-on interventions to mobile apps. This meta-analysis also included gray literature such as conference abstracts, and the funnel plots showed good symmetry, excluding the possibility of publication bias. In contrast, the findings of this meta-analysis should be treated with caution because of the vast heterogeneity in the studies that would limit real-life applicability. Moreover, one-third of the articles had a high risk of bias; however, this bias could not have been avoided because of the nature of the app and its effect on blinding. Although some of the studies used commercial apps, the study team developed most of them. Another limitation is the heterogeneous behavioral component that ranged from simple booklets to in-person meetings and telephone calls. Finally, it is essential to note that the weight loss outcome was measured in kilograms rather than as a percentage of weight loss from baseline.

Conclusions

Mobile phone apps have a role in weight loss management and result in modest weight loss compared with active control. Combining a mobile app, tracker, and human-delivered behavioral component led to the highest degree of weight loss at 6 months. Further research should use artificial intelligence to replace the human-delivered behavioral component for better mobile app use scalability.

Authors' Contributions

JA was responsible for the research idea, design, data collection, analysis, and manuscript writing. HI was responsible for the research design, data collection, and manuscript writing. NA was responsible for the data collection, analysis, and manuscript writing. AE was responsible for the data collection and manuscript writing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Examples of the search strategy with keywords.
[DOCX File , 12 KB - mhealth_v10i4e35479_app1.docx]

Multimedia Appendix 2
Summary of the excluded articles and the reasons for exclusion.
[DOCX File , 118 KB - mhealth_v10i4e35479_app2.docx]

Multimedia Appendix 3
Funnel plots.
[DOCX File , 36 KB - mhealth_v10i4e35479_app3.docx]

Multimedia Appendix 4
Summary of the components of the intervention and control arms of the included studies.
[DOCX File , 27 KB - mhealth_v10i4e35479_app4.docx]

Multimedia Appendix 5
Summary of the adherence measures.
[DOCX File , 17 KB - mhealth_v10i4e35479_app5.docx]

References

1. Hartmann-Boyce J, Johns DJ, Jebb SA, Aveyard P, Behavioural Weight Management Review Group. Effect of behavioural techniques and delivery mode on effectiveness of weight management: systematic review, meta-analysis and meta-regression. Obes Rev 2014 Jul;15(7):598-609 [FREE Full text] [doi: 10.1111/obr.12165] [Medline: 24636238]
2. Pagoto S, Schneider K, Jojic M, DeBiasse M, Mann D. Evidence-based strategies in weight-loss mobile apps. Am J Prev Med 2013 Nov;45(5):576-582. [doi: 10.1016/j.amepre.2013.04.025] [Medline: 24139770]
3. Azar KM, Lesser LI, Laing BY, Stephens J, Aurora MS, Burke LE, et al. Mobile applications for weight management: theory-based content analysis. Am J Prev Med 2013 Nov;45(5):583-589. [doi: 10.1016/j.amepre.2013.07.005] [Medline: 24139771]
4. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
5. Podina IR, Fodor LA, Cosmoiu A, Boian R. An evidence-based gamified mHealth intervention for overweight young adults with maladaptive eating habits: study protocol for a randomized controlled trial. Trials 2017 Dec 12;18(1):592 [FREE Full text] [doi: 10.1186/s13063-017-2340-6] [Medline: 29233169]
6. Tang J, Abraham C, Stamp E, Greaves C. How can weight-loss app designers' best engage and support users? A qualitative investigation. Br J Health Psychol 2015 Feb;20(1):151-171. [doi: 10.1111/bjhp.12114] [Medline: 25130682]
7. Podina IR, Fodor LA. Critical review and meta-analysis of multicomponent behavioral e-health interventions for weight loss. Health Psychol 2018 Jun;37(6):501-515. [doi: 10.1037/hea0000623] [Medline: 29733617]
8. Wharton CM, Johnston CS, Cunningham BK, Sterner D. Dietary self-monitoring, but not dietary quality, improves with use of smartphone app technology in an 8-week weight loss trial. J Nutr Educ Behav 2014;46(5):440-444. [doi: 10.1016/j.jneb.2014.04.291] [Medline: 25220777]
9. Lieffers JR, Arocha JF, Grindrod K, Hanning RM. Experiences and perceptions of adults accessing publicly available nutrition behavior-change mobile apps for weight management. J Acad Nutr Diet 2018 Feb;118(2):229-39.e3. [doi: 10.1016/j.jand.2017.04.015] [Medline: 28625662]
10. Yu Z, Sealey-Potts C, Rodriguez J. Dietary self-monitoring in weight management: current evidence on efficacy and adherence. J Acad Nutr Diet 2015 Dec;115(12):1931-1938. [doi: 10.1016/j.jand.2015.04.005] [Medline: 26028176]
11. Cavero-Redondo I, Martinez-Vizcaino V, Fernandez-Rodriguez R, Saz-Lara A, Pascual-Morena C, Álvarez-Bueno C. Effect of behavioral weight management interventions using lifestyle mHealth self-monitoring on weight loss: a systematic review and meta-analysis. Nutrients 2020 Jul 03;12(7):1977 [FREE Full text] [doi: 10.3390/nu12071977] [Medline: 32635174]
12. Carter MC, Burley VJ, Nykjaer C, Cade JE. Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. J Med Internet Res 2013 Apr 15;15(4):e32 [FREE Full text] [doi: 10.2196/jmir.2283] [Medline: 23587561]

13. Flores Mateo G, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile phone apps to promote weight loss and increase physical activity: a systematic review and meta-analysis. J Med Internet Res 2015 Nov 10;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836] [Medline: 26554314]
14. Khokhar B, Jones J, Ronksley PE, Armstrong MJ, Caird J, Rabi D. Effectiveness of mobile electronic devices in weight loss among overweight and obese populations: a systematic review and meta-analysis. BMC Obes 2014;1:22 [FREE Full text] [doi: 10.1186/s40608-014-0022-4] [Medline: 26217509]
15. Hutchesson MJ, Tan CY, Morgan P, Callister R, Collins C. Enhancement of self-monitoring in a web-based weight loss program by extra individualized feedback and reminders: randomized trial. J Med Internet Res 2016 Apr 12;18(4):e82 [FREE Full text] [doi: 10.2196/jmir.4100] [Medline: 27072817]
16. Liu F, Kong X, Cao J, Chen S, Li C, Huang J, et al. Mobile phone intervention and weight loss among overweight and obese adults: a meta-analysis of randomized controlled trials. Am J Epidemiol 2015 Mar 01;181(5):337-348 [FREE Full text] [doi: 10.1093/aje/kwu260] [Medline: 25673817]
17. Cai X, Qiu S, Luo D, Wang L, Lu Y, Li M. Mobile application interventions and weight loss in type 2 diabetes: a meta-analysis. Obesity (Silver Spring) 2020 Mar;28(3):502-509. [doi: 10.1002/oby.22715] [Medline: 31965748]
18. Islam MM, Poly TN, Walther BA, Jack Li YC. Use of mobile phone app interventions to promote weight loss: meta-analysis. JMIR Mhealth Uhealth 2020 Jul 22;8(7):e17039 [FREE Full text] [doi: 10.2196/17039] [Medline: 32706724]
19. Bennett GG, Steinberg D, Askew S, Levine E, Foley P, Batch BC, et al. Effectiveness of an app and provider counseling for obesity treatment in primary care. Am J Prev Med 2018 Dec;55(6):777-786 [FREE Full text] [doi: 10.1016/j.amepre.2018.07.005] [Medline: 30361140]
20. Arens JH, Hauth W, Weissmann J. Novel app- and web-supported diabetes prevention program to promote weight reduction, physical activity, and a healthier lifestyle: observation of the clinical application. J Diabetes Sci Technol 2018 Jul;12(4):831-838 [FREE Full text] [doi: 10.1177/1932296818768621] [Medline: 29584454]
21. Everett E, Kane B, Yoo A, Dobs A, Mathioudakis N. A novel approach for fully automated, personalized health coaching for adults with prediabetes: pilot clinical trial. J Med Internet Res 2018 Feb 27;20(2):e72 [FREE Full text] [doi: 10.2196/jmir.9723] [Medline: 29487046]
22. Lyzwinski LN, Caffery LJ, Bambling M, Edirippulige S. Consumer perspectives on mHealth for weight loss: a review of qualitative studies. J Telemed Telecare 2018 May;24(4):290-302. [doi: 10.1177/1357633X17692722] [Medline: 28181859]
23. Semper HM, Povey R, Clark-Carter D. A systematic review of the effectiveness of smartphone applications that encourage dietary self-regulatory strategies for weight loss in overweight and obese adults. Obes Rev 2016 Sep;17(9):895-906. [doi: 10.1111/obr.12428] [Medline: 27192162]
24. Webb VL, Wadden TA. Intensive lifestyle intervention for obesity: principles, practices, and results. Gastroenterology 2017 May;152(7):1752-1764. [doi: 10.1053/j.gastro.2017.01.045] [Medline: 28192109]
25. Levine DM, Savarimuthu S, Squires A, Nicholson J, Jay M. Technology-assisted weight loss interventions in primary care: a systematic review. J Gen Intern Med 2015 Jan;30(1):107-117 [FREE Full text] [doi: 10.1007/s11606-014-2987-6] [Medline: 25134692]
26. Thompson-Felty C, Johnston CS. Adherence to diet applications using a smartphone was associated with weight loss in healthy overweight adults irrespective of the application. J Diabetes Sci Technol 2017 Jan;11(1):184-185 [FREE Full text] [doi: 10.1177/1932296816656209] [Medline: 27334302]
27. Ross KM, Wing RR. Impact of newer self-monitoring technology and brief phone-based intervention on weight loss: a randomized pilot study. Obesity (Silver Spring) 2016 Aug;24(8):1653-1659 [FREE Full text] [doi: 10.1002/oby.21536] [Medline: 27367614]
28. Spring B, Pellegrini CA, Pfammatter A, Duncan JM, Pictor A, McFadden HG, et al. Effects of an abbreviated obesity intervention supported by mobile technology: the ENGAGED randomized clinical trial. Obesity (Silver Spring) 2017 Jul;25(7):1191-1198 [FREE Full text] [doi: 10.1002/oby.21842] [Medline: 28494136]
29. Svetkey LP, Batch BC, Lin PH, Intille SS, Corsino L, Tyson CC, et al. Cell phone intervention for you (CITY): a randomized, controlled trial of behavioral weight loss intervention for young adults using mobile technology. Obesity (Silver Spring) 2015 Nov;23(11):2133-2141 [FREE Full text] [doi: 10.1002/oby.21226] [Medline: 26530929]
30. Lyzwinski L. A systematic review and meta-analysis of mobile devices and weight loss with an intervention content analysis. J Pers Med 2014 Jun 30;4(3):311-385 [FREE Full text] [doi: 10.3390/jpm4030311] [Medline: 25563356]
31. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
32. Bender MS, Cooper B, Arai S. A feasible and effective mobile health weight loss lifestyle intervention for Filipinos with type 2 diabetes. Circulation 2017 Mar 07;135(suppl_1):MP006. [doi: 10.1161/circ.135.suppl_1.mp006]
33. Fukuoka Y, Gay CL, Joiner KL, Vittinghoff E. A novel diabetes prevention intervention using a mobile app: a randomized controlled trial with overweight adults at risk. Am J Prev Med 2015 Aug;49(2):223-237 [FREE Full text] [doi: 10.1016/j.amepre.2015.01.003] [Medline: 26033349]

34. Whitelock V, Kersbergen I, Higgs S, Aveyard P, Halford JC, Robinson E. A smartphone based attentive eating intervention for energy intake and weight loss: results from a randomised controlled trial. BMC Public Health 2019 May 21;19(1):611 [FREE Full text] [doi: 10.1186/s12889-019-6923-x] [Medline: 31113400]
35. Vaz CL, Suthar AG, Pousti BT, Aye SM, Williams KJ, Zhao H. A smartphone app-based lifestyle intervention promotes weight loss-results of a prospective, randomized, controlled clinical trial (RCT). Diabetes 2018 Jul 1;67(Supplement_1):A543. [doi: 10.2337/db18-2070-P]
36. Rogers RJ, Lang W, Barone Gibbs B, Davis KK, Burke LE, Kovacs SJ, et al. Applying a technology-based system for weight loss in adults with obesity. Obes Sci Pract 2016 Mar;2(1):3-12 [FREE Full text] [doi: 10.1002/osp4.18] [Medline: 27812375]
37. Thomas JG, Bond DS, Raynor HA, Papandonatos GD, Wing RR. Comparison of smartphone-based behavioral obesity treatment with gold standard group treatment and control: a randomized trial. Obesity (Silver Spring) 2019 Apr;27(4):572-580 [FREE Full text] [doi: 10.1002/oby.22410] [Medline: 30779333]
38. Brindal E, Hendrie G, Freyne J, Coombe M, Berkovsky S, Noakes M. Design and pilot results of a mobile phone weight-loss application for women starting a meal replacement programme. J Telemed Telecare 2013 Apr;19(3):166-174. [doi: 10.1177/1357633X13479702] [Medline: 23520213]
39. Laing BY, Mangione CM, Tseng CH, Leng M, Vaisberg E, Mahida M, et al. Effectiveness of a smartphone application for weight loss compared with usual care in overweight primary care patients: a randomized, controlled trial. Ann Intern Med 2014 Nov 18;161(10 Suppl):S5-12 [FREE Full text] [doi: 10.7326/M13-3005] [Medline: 25402403]
40. Shin DW, Yun JM, Shin JH, Kwon H, Min HY, Joh HK, et al. Enhancing physical activity and reducing obesity through smartcare and financial incentives: a pilot randomized trial. Obesity (Silver Spring) 2017 Feb;25(2):302-310. [doi: 10.1002/oby.21731] [Medline: 28063226]
41. Gilmore LA, Klempel MC, Martin CK, Myers CA, Burton JH, Sutton EF, et al. Personalized mobile health intervention for health and weight loss in postpartum women receiving women, infants, and children benefit: a randomized controlled pilot study. J Womens Health (Larchmt) 2017 Jul;26(7):719-727 [FREE Full text] [doi: 10.1089/jwh.2016.5947] [Medline: 28338403]
42. Tanaka K, Sasai H, Wakaba K, Murakami S, Ueda M, Yamagata F, et al. Professional dietary coaching within a group chat using a smartphone application for weight loss: a randomized controlled trial. J Multidiscip Healthc 2018;11:339-347 [FREE Full text] [doi: 10.2147/JMDH.S165422] [Medline: 30038502]
43. Allen JK, Stephens J, Dennison Himmelfarb CR, Stewart KJ, Hauck S. Randomized controlled pilot study testing use of smartphone technology for obesity treatment. J Obes 2013;2013:151597 [FREE Full text] [doi: 10.1155/2013/151597] [Medline: 24392223]
44. Stephens JD, Yager AM, Allen J. Smartphone technology and text messaging for weight loss in young adults: a randomized controlled trial. J Cardiovasc Nurs 2017;32(1):39-46 [FREE Full text] [doi: 10.1097/JCN.0000000000000307] [Medline: 26646593]
45. Hales S, Turner-McGrievy GM, Wilcox S, Fahim A, Davis RE, Huhns M, et al. Social networks for improving healthy weight loss behaviors for overweight and obese adults: a randomized clinical trial of the social pounds off digitally (Social POD) mobile app. Int J Med Inform 2016 Oct;94:81-90. [doi: 10.1016/j.ijmedinf.2016.07.003] [Medline: 27573315]
46. Hartman SJ, Nelson SH, Cadmus-Bertram LA, Patterson RE, Parker BA, Pierce JP. Technology- and phone-based weight loss intervention: pilot RCT in women at elevated breast cancer risk. Am J Prev Med 2016 Nov;51(5):714-721 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.024] [Medline: 27593420]
47. Haufe S, Kerling A, Protte G, Bayerle P, Stenner HT, Rolff S, et al. Telemonitoring-supported exercise training, metabolic syndrome severity, and work ability in company employees: a randomised controlled trial. Lancet Public Health 2019 Jul;4(7):e343-e352 [FREE Full text] [doi: 10.1016/S2468-2667(19)30075-1] [Medline: 31204284]
48. Turner-McGrievy GM, Wilcox S, Boutté A, Hutto BE, Singletary C, Muth ER, et al. The dietary intervention to enhance tracking with mobile devices (DIET mobile) study: a 6-month randomized weight loss trial. Obesity (Silver Spring) 2017 Aug;25(8):1336-1342 [FREE Full text] [doi: 10.1002/oby.21889] [Medline: 28600833]
49. Jospe MR, Roy M, Brown RC, Williams SM, Osborne HR, Meredith-Jones KA, et al. The effect of different types of monitoring strategies on weight loss: a randomized controlled trial. Obesity (Silver Spring) 2017 Sep;25(9):1490-1498. [doi: 10.1002/oby.21898] [Medline: 28703448]
50. Burke LE, Zheng Y, Ma Q, Mancino J, Loar I, Music E, et al. The SMARTER pilot study: testing feasibility of real-time feedback for dietary self-monitoring. Prev Med Rep 2017 Jun;6:278-285 [FREE Full text] [doi: 10.1016/j.pmedr.2017.03.017] [Medline: 28409090]
51. Lee JH, Shim J, Yeom H, Lee SJ, Kim HC. Abstract MP69: The weight loss by using the smartphone application for people with metabolic abnormalities: a randomized controlled trial. Circulation 2019 Mar 05;139(Suppl_1):MP69. [doi: 10.1161/circ.139.suppl_1.mp69]
52. Turner-McGrievy G, Tate D. Tweets, apps, and pods: results of the 6-month mobile pounds off digitally (mobile POD) randomized weight-loss intervention among adults. J Med Internet Res 2011 Dec 20;13(4):e120 [FREE Full text] [doi: 10.2196/jmir.1841] [Medline: 22186428]

53. Monroe CM, Geraci M, Larsen CA, West DS. Feasibility and efficacy of a novel technology-based approach to harness social networks for weight loss: the NETworks pilot randomized controlled trial. Obes Sci Pract 2019 Aug;5(4):354-365 [FREE Full text] [doi: 10.1002/osp4.352] [Medline: 31452920]
54. Choi B, Dhawan T, Metzger K, Marshall L, Akbar A, Jain T, et al. Intervention to alter dietary consumption to a Mediterranean diet in an American cardiology patient population: a randomized controlled trial of coaching delivered via smartphone app versus traditional counseling. J Am Coll Cardiol 2018 Mar;71(11):A1741. [doi: 10.1016/s0735-1097(18)32282-4]
55. Evangelista L, Cacciata M, Lombardo D, Stromberg A. Reduced body weight, improved physical activity and healthier eating with fitness intensive therapy (Get FIT) vs. Get FIT+: a feasibility study. J Heart Lung Transplant 2018 Apr;37(4):S118. [doi: 10.1016/j.healun.2018.01.281]
56. Kurtzman GW, Day SC, Small DS, Lynch M, Zhu J, Wang W, et al. Social incentives and gamification to promote weight loss: the LOSE IT randomized, controlled trial. J Gen Intern Med 2018 Oct;33(10):1669-1675 [FREE Full text] [doi: 10.1007/s11606-018-4552-1] [Medline: 30003481]
57. Duncan MJ, Fenton S, Brown WJ, Collins CE, Glozier N, Kolt GS, et al. Efficacy of a multi-component m-health weight-loss intervention in overweight and obese adults: a randomised controlled trial. Int J Environ Res Public Health 2020 Aug 26;17(17):6200 [FREE Full text] [doi: 10.3390/ijerph17176200] [Medline: 32859100]
58. Lim SL, Ong KW, Johal J, Han CY, Yap QV, Chan YH, et al. Effect of a smartphone app on weight change and metabolic outcomes in Asian adults with type 2 diabetes: a randomized clinical trial. JAMA Netw Open 2021 Jun 01;4(6):e2112417 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.12417] [Medline: 34081137]
59. Ahn JS, Lee H, Kim J, Park H, Kim DW, Lee JE. Use of a smartphone app for weight loss versus a paper-based dietary diary in overweight adults: randomized controlled trial. JMIR Mhealth Uhealth 2020 Jul 31;8(7):e14013 [FREE Full text] [doi: 10.2196/14013] [Medline: 32735225]
60. Lugones-Sanchez C, Sanchez-Calavera MA, Repiso-Gento I, Adalia EG, Ramirez-Manent JI, Agudo-Conde C, EVIDENT 3 Investigators. Effectiveness of an mHealth intervention combining a smartphone app and smart band on body composition in an overweight and obese population: randomized controlled trial (EVIDENT 3 study). JMIR Mhealth Uhealth 2020 Nov 26;8(11):e21771 [FREE Full text] [doi: 10.2196/21771] [Medline: 33242020]
61. Nyström CD, Sandin S, Henriksson P, Henriksson H, Trolle-Lagerros Y, Larsson C, et al. Mobile-based intervention intended to stop obesity in preschool-aged children: the MINISTOP randomized controlled trial. Am J Clin Nutr 2017 Jun;105(6):1327-1335. [doi: 10.3945/ajcn.116.150995] [Medline: 28446496]
62. Rivera J, McPherson A, Hamilton J, Birken C, Coons M, Iyer S, et al. Mobile apps for weight management: a scoping review. JMIR Mhealth Uhealth 2016 Jul 26;4(3):e87 [FREE Full text] [doi: 10.2196/mhealth.5115] [Medline: 27460502]
63. Stein N, Brooks K. A fully automated conversational artificial intelligence for weight loss: longitudinal observational study among overweight and obese adults. JMIR Diabetes 2017 Nov 01;2(2):e28 [FREE Full text] [doi: 10.2196/diabetes.8590] [Medline: 30291087]
64. Pereira J, Díaz Ó. Using health chatbots for behavior change: a mapping study. J Med Syst 2019 Apr 04;43(5):135. [doi: 10.1007/s10916-019-1237-1] [Medline: 30949846]
65. Zhang J, Oh YJ, Lange P, Yu Z, Fukuoka Y. Artificial intelligence chatbot behavior change model for designing artificial intelligence chatbots to promote physical activity and a healthy diet: viewpoint. J Med Internet Res 2020 Sep 30;22(9):e22845 [FREE Full text] [doi: 10.2196/22845] [Medline: 32996892]
66. Huang CY, Yang MC. Empirical investigation of factors influencing consumer intention to use an artificial intelligence-powered mobile application for weight loss and health management. Telemed J E Health 2020 Oct;26(10):1240-1251. [doi: 10.1089/tmj.2019.0182] [Medline: 31971883]
67. Varkevisser RD, van Stralen MM, Kroeze W, Ket JC, Steenhuis IH. Determinants of weight loss maintenance: a systematic review. Obes Rev 2019 Feb;20(2):171-211 [FREE Full text] [doi: 10.1111/obr.12772] [Medline: 30324651]
68. Sorgente A, Pietrabissa G, Manzoni GM, Re F, Simpson S, Perona S, et al. Web-based interventions for weight loss or weight loss maintenance in overweight and obese people: a systematic review of systematic reviews. J Med Internet Res 2017 Jun 26;19(6):e229 [FREE Full text] [doi: 10.2196/jmir.6972] [Medline: 28652225]
69. Peterson ND, Middleton KR, Nackers LM, Medina KE, Milsom VA, Perri MG. Dietary self-monitoring and long-term success with weight management. Obesity (Silver Spring) 2014 Sep;22(9):1962-1967 [FREE Full text] [doi: 10.1002/oby.20807] [Medline: 24931055]
70. Krukowski RA, Harvey-Berino J, Bursac Z, Ashikaga T, West DS. Patterns of success: online self-monitoring in a web-based behavioral weight control program. Health Psychol 2013 Feb;32(2):164-170 [FREE Full text] [doi: 10.1037/a0028135] [Medline: 22545978]
71. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
72. Payne JE, Turk MT, Kalarchian MA, Pellegrini CA. Defining adherence to dietary self-monitoring using a mobile app: a narrative review. J Acad Nutr Diet 2018 Nov;118(11):2094-2119. [doi: 10.1016/j.jand.2018.05.011] [Medline: 30115555]

Abbreviations

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

Edited by L Buis; submitted 06.12.21; peer-reviewed by G Papandonatos, B Kang; comments to author 25.01.22; revised version received 09.02.22; accepted 11.03.22; published 08.04.22.

Please cite as:
Antoun J, Itani H, Alarab N, Elsehmawy A
The Effectiveness of Combining Nonmobile Interventions With the Use of Smartphone Apps With Various Features for Weight Loss: Systematic Review and Meta-analysis
JMIR Mhealth Uhealth 2022;10(4):e35479
URL: https://mhealth.jmir.org/2022/4/e35479
doi:10.2196/35479
PMID:35394443

Review

The Effectiveness of Wearable Devices as Physical Activity Interventions for Preventing and Treating Obesity in Children and Adolescents: Systematic Review and Meta-analysis

Wentao Wang[1], MEd; Jing Cheng[1], MEd; Weijun Song[1], MEd; Yi Shen[1], MEd

Department of Basic Education, Zhejiang Tongji Vocational College of Science and Technology, Hangzhou, China

Corresponding Author:
Wentao Wang, MEd
Department of Basic Education
Zhejiang Tongji Vocational College of Science and Technology
Number 418, Gengwen Road, Xiaoshan District
Hangzhou, 311231
China
Phone: 86 15068156990
Email: 371457623@qq.com

Abstract

Background: The prevalence of obesity in children and adolescents remains a global public health issue. Wearable devices may offer new opportunities for prevention and intervention in obesity. Previous systematic reviews have only examined the effect of the wearable device interventions on preventing and treating obesity in adults. However, no systematic review has provided an evaluation of wearable devices as physical activity interventions for preventing and treating obesity in children and adolescents.

Objective: The purpose of this review and meta-analysis was to evaluate the effectiveness of wearable devices as physical activity interventions on obesity-related anthropometric outcomes in children and adolescents.

Methods: Research articles retrieved from PubMed, EMBASE, Cochrane Library, Scopus, and EBSCO from inception to February 1, 2021, were reviewed. The search was designed to identify studies utilizing wearable devices for preventing and treating obesity in children and adolescents. The included studies were evaluated for risk of bias following the Cochrane recommendation. Meta-analyses were conducted to evaluate the effectiveness of wearable devices as physical activity interventions on body weight, body fat, BMI z-score (BMI-Z), BMI, and waist circumference. Subgroup analyses were performed to determine whether the characteristics of the interventions had an impact on the effect size.

Results: A total of 12 randomized controlled trials (3227 participants) were selected for meta-analysis. Compared with the control group, wearable device interventions had statistically significant beneficial effects on BMI (mean difference [MD] –0.23; 95% CI –0.43 to –0.03; P=.03; I^2=2%), BMI-Z (MD –0.07; 95% CI –0.13 to –0.01; P=.01; I^2=81%), body weight (MD –1.08; 95% CI –2.16 to –0.00; P=.05; I^2=58%), and body fat (MD –0.72; 95% CI –1.19 to –0.25; P=.003; I^2=5%). However, no statistically significant effect was found on waist circumference (MD 0.55; 95% CI –0.21 to 1.32; P=.16; I^2=0%). The subgroup analysis showed that for participants with overweight or obesity (MD –0.75; 95% CI –1.18 to –0.31; P<.01; I^2=0%), in the short-term (MD –0.62; 95% CI –1.03 to –0.21; P<.01; I^2=0%), wearable-based interventions (MD –0.56; 95% CI –0.95 to –0.18; P<.01; I^2=0%) generally resulted in greater intervention effect size on BMI.

Conclusions: Evidence from this meta-analysis shows that wearable devices as physical activity interventions may be useful for preventing and treating obesity in children and adolescents. Future research is needed to identify the most effective physical activity indicators of wearable devices to prevent and treat obesity in children and adolescents.

(JMIR Mhealth Uhealth 2022;10(4):e32435) doi:10.2196/32435

KEYWORDS

wearable devices; obesity; children; adolescents; meta-analysis

Introduction

With the development of society and technology, human lifestyles have undergone tremendous changes. The worldwide prevalence of obesity has risen rapidly since 1975. In 2016, more than 340 million children and adolescents aged 5-19 were overweight or obese [1]. As a global public health issue, obesity might cause a number of serious health conditions, such as high blood pressure, nonalcoholic fatty liver disease, abnormal lipid metabolism, and psychosocial problems [2-5]. Therefore, effective interventions for preventing and treating obesity in children and adolescents are urgently needed.

At present, regular physical activity seems to be one of the effective means for the prevention of and intervention in obesity among children and adolescents and has been discussed in many studies [6-8]. Traditional physical activity intervention methods (such as school group physical activity interventions [9] and face-to-face interventions with health professionals [10]) require significant effort and costs. However, wearable devices may provide an alternative means of addressing obesity in children and adolescents. Wearable devices, such as pedometers, sports bracelets, sports watches, and accelerometers, can offer easy and effective ways to collect physical activity data (steps, heart rate, energy expenditure, physical activity, and physical activity time of different intensity) and allow users to monitor their data [11,12]. These quantitative data can stimulate the user's motivation for physical activity, increase physical activity time, and increase energy expenditure, thus enabling weight loss [13-18].

A considerable body of work related to the use of wearable devices to prevent and treat obesity has been already published [19-23]. Most previous reviews have focused on adults, in which it was demonstrated that wearable devices can achieve a significant effect size in reducing the BMI of adults with obesity (β=−1.57; *P*<.001) or adults with chronic diseases (β=−1.30; *P*<.001). However, evidence on the effectiveness of wearable device interventions remains inconclusive. Furthermore, wearable devices have no significant effect size on adults with normal weight (β=−0.49; *P*=.07) [17,24-26]. The characteristics and cognitive abilities of children and adolescents are different from those of adults, and emerging technologies have great appeal to children and adolescents [27,28]. Children and adolescents are in a sensitive period of growth and development. Compared with diet control, children and adolescents are more suited to perform increased physical activity for preventing and treating obesity [29,30]. Wearable devices have been proven to be valid and accurate [31-35], providing consistent feedback and inducing behavior changes in individuals, which result in increased physical activity [36-41]. Therefore, there is a need to explore the effectiveness of wearable devices as physical activity interventions for preventing and treating obesity in this population.

To the best of our knowledge, there are no reviews evaluating wearable devices as physical activity interventions for preventing and treating obesity in children and adolescents. It is thus necessary to explore the effectiveness of wearable devices as physical activity interventions to prevent and treat obesity specifically in these populations. The results may contribute to public health guidance on the use of wearable devices for addressing obesity in children and adolescents. Therefore, the objectives of this review and meta-analysis are to (1) evaluate the effectiveness of wearable devices as physical activity interventions on obesity-related anthropometric outcomes in children and adolescents and (2) determine whether the characteristics of the interventions had an impact on the effect size through subgroup analyses.

Methods

Study Design

This systematic review and meta-analysis is reported in accordance with the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) guidelines (Multimedia Appendix 1) [42].

Search Strategy

The following 5 international electronic databases were searched to discover studies on the use of wearable devices for preventing and treating obesity in children and adolescents: PubMed, EMBASE, Cochrane Library, Scopus, and EBSCO. The time span was set from the inception of each database to February 1, 2021. Search strategies were adapted according to the requirements of each database (see Multimedia Appendix 2 for the complete PubMed search strategy). At the same time, relevant articles were also found by checking the references of the included articles and previous systematic reviews.

Inclusion and Exclusion Criteria

The inclusion criteria included studies with the following characteristics:

1. Population: The participants were children and adolescents aged 6-18 years.
2. Interventions: The intervention groups involved the use of wearable devices to promote physical activity. Wearable devices need to be worn on the user's body, and use accelerometers or sensors to track the wearer's physical activity or physiological data, such as wristbands, pedometers, smartwatches.
3. Outcomes: The outcome featured obesity-related anthropometric indicators, such as BMI, BMI z-score (BMI-Z), body weight, or body fat.
4. The experimental design was a randomized controlled trial (RCT).

The exclusion criteria were as follows:

1. The participants were aged <6 or >18 years.
2. The intervention did not involve wearable devices (eg, smartphones, video games, and social media) or their use was not related to promoting physical activity (eg, monitor food consumption and strengthen communication and guidance).
3. The primary outcomes were not obesity-related anthropometric indicators (eg, quality of life, food consumption, and psychological state).
4. The experimental design was not an RCT.
5. The articles were meta-analyses or systematic reviews.

The literature screening was first conducted independently by 2 authors (WS and YS), according to the inclusion and exclusion criteria described above. Then the 2 researchers cross-checked the included literature. Following this, documents for which eligibility was unclear were selected according to consensus of a third author (WW).

Data Extraction and Risk of Bias Assessment

Two reviewers (WS and YS) extracted data independently from each included study. The extracted content included author, region, year of publication, clinical research design, research object, sample size, population characteristics, intervention method, intervention period, and outcome indicators.

Two reviewers (JC and YS) independently evaluated each study for risk of bias following the Cochrane recommendations. Each criterion was scored as having a low, unclear, or high level of risk. The evaluation content included (1) allocation concealment, (2) random sequence generation, (3) blinding of the outcome, (4) blinding of participants and personnel, (5) selective reporting, (6) incomplete outcome data, and (7) other bias [43].

Statistical Analysis

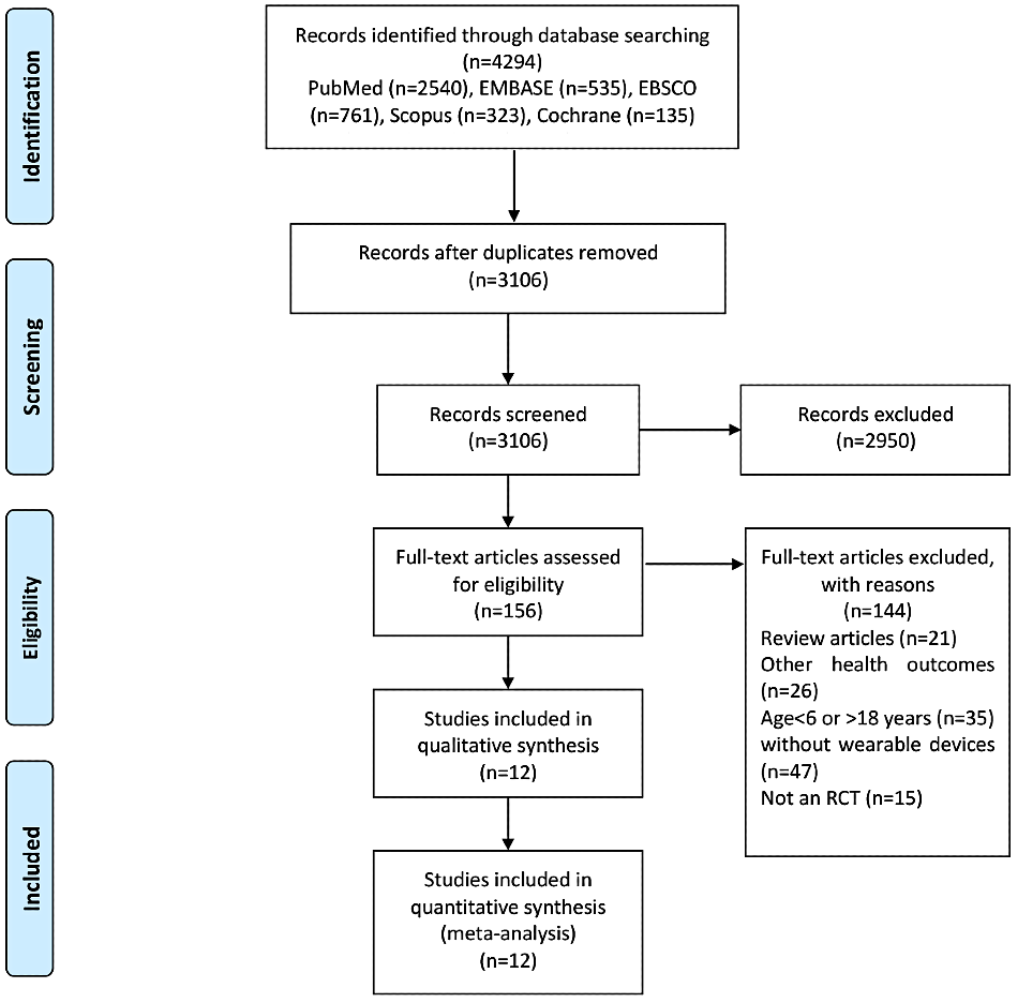
RevMan5.3 software (International Cochrane Collaboration) was used for the meta-analysis. The mean difference (MD) and 95% CI were used to represent continuous variables. First, the included studies were tested for heterogeneity at a level of OR of 0.05. When $I^2 \leq 50\%$, there is no statistically significant heterogeneity and the fixed effects model should be used; when $I^2 > 50\%$, there is a statistical heterogeneity between the studies and the random effects model should be used. For studies with heterogeneity, a sensitivity or subgroup analysis could be carried out; for studies with clinical heterogeneity and methodological heterogeneity, a meta-analysis of outcome indicators was abandoned, and a general statistical description was used.

Results

Study Selection

The flowchart in Figure 1 outlines the process for selecting articles for inclusion. A total of 4294 related articles were identified from 5 databases. A total of 1188 articles were eliminated as duplicates, and a further 2950 records were eliminated after screening of titles and abstracts. The remaining 156 articles were reviewed in detail and evaluated based on their full text. A further 26 articles were eliminated because the outcomes did not include obesity indicators, 35 owing to their irrelevance to children or adolescents, 47 for not featuring wearable devices, and 15 because their experimental design was not an RCT. Finally, 12 articles were selected for this meta-analysis.

Figure 1. The selection process for the systematic review and meta-analysis. RCT: randomized controlled trial.



Risk-of-Bias Assessment

The risk-of-bias assessment results are presented in Multimedia Appendix 3. All studies were assessed as having a low risk of bias in terms of blinding of the outcome and selective reporting. In terms of random sequence generation, 11 studies had a low risk of bias, and only 1 study had a high risk of bias [44]. In the allocation concealment, 10 studies had a high risk of bias or were unclear, and 2 studies had a low risk of bias [45,46]. In terms of blinding of participants and personnel, all studies had a high risk of bias, due to the nature of the intervention. In the incomplete outcome data, 4 studies had a high risk of bias [46-49], all of which had a high rate of attrition (>25%). As many as 2 studies had a high risk of other bias [47,49], because the baseline of the intervention group and the control group was significantly different (*P*=.02).

Characteristics of Included Studies

The characteristics of the included studies are presented in Multimedia Appendix 4. A many as 4 studies were carried out in the United States [48-51], 4 in Australia [44,46,52,53], 1 in Finland [23], 1 in Italy [47], 1 in Germany [54], and 1 in Singapore [45]. These included studies were published between 2011 and 2021 and involved 3227 participants. The average age of participants was 13.2 years and ranged from 6 to 18 years. The dropout rate was studied for each intervention and ranged from 3.9% (4/102) to 52.0% (25/48), with the average being 19.27% (622/3227). Intervention duration ranged from 2.5 to 18 months, with the average being 6.2 months. Among these 12 studies, 8 intervention targets were people of normal weight [23,44-46,50,52-54] and 4 were those with overweight or obesity [47-49,51]. A total of 7 studies used pedometers [44-46,49,52-54] and the other 5 used wristband activity trackers [23,47,48,50,51]. The most common outcome index was BMI [23,44-46,48-53], followed by BMI-Z [46,47,49,51-53], body fat [23,44,50,52-54], body weight [23,47-49,52], and waist circumference [23,44,46,48,52]. There were 2 wearable device intervention groups in the same studies [49,50]. The data from each intervention group were included in the meta-analysis as independent samples.

Effects of Intervention

BMI

A total of 10 studies explored the effects of wearable devices as physical activity interventions on the BMI of children and adolescents [23,44-46,48-53]. There were statistically significant decreases in BMI between the group with wearable device interventions and the control group (MD –0.23; 95% CI –0.43 to –0.03; *P*=.03); heterogeneity was low and insignificant (I^2=2%; *P*=.43; Figure 2). In this analysis, the study conducted by Smith et al [44] had the greatest proportion (59.6%).

A subgroup analysis found that, compared with participants with normal weight (MD –0.09; 95% CI –0.32 to 0.14; *P*=.46; I^2=0%; Figure 3), participants who were overweight or obese had a significantly greater intervention effect size on BMI (MD –0.75; 95% CI –1.18 to –0.31; *P*<.01; I^2=0%). Another subgroup analysis showed that interventions with an estimated duration of ≤4 months (MD –0.62; 95% CI –1.03 to –0.21; *P*<.01; I^2=0%; Figure 4) had a significantly greater intervention effect size on BMI than those with an estimated duration of >4 months (MD –0.10; 95% CI –0.34 to 0.13; *P*=.39; I^2=0%). Further subgroup analysis demonstrated that, compared with the multifaceted intervention program (MD –0.10; 95% CI –0.34 to 0.13; *P*=.40; I^2=0%; Figure 5), the wearable-based intervention programs caused a significantly greater decrease in BMI (MD –0.56; 95% CI –0.95 to –0.18; *P*<.01; I^2=0%).

Figure 2. Forest plot of the effect of the wearable device interventions on BMI.

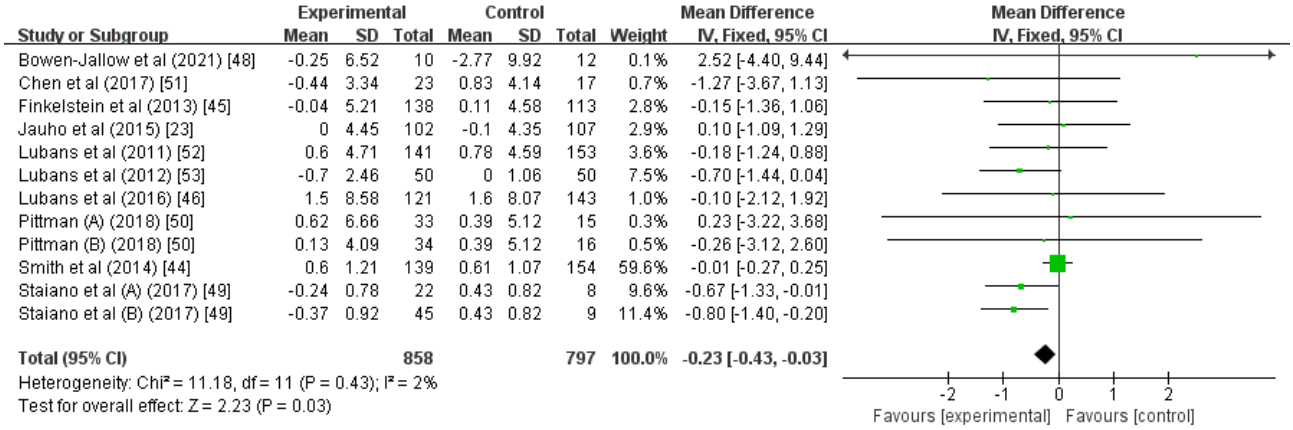


Figure 3. Forest plot of the participant characteristics subgroup analysis.

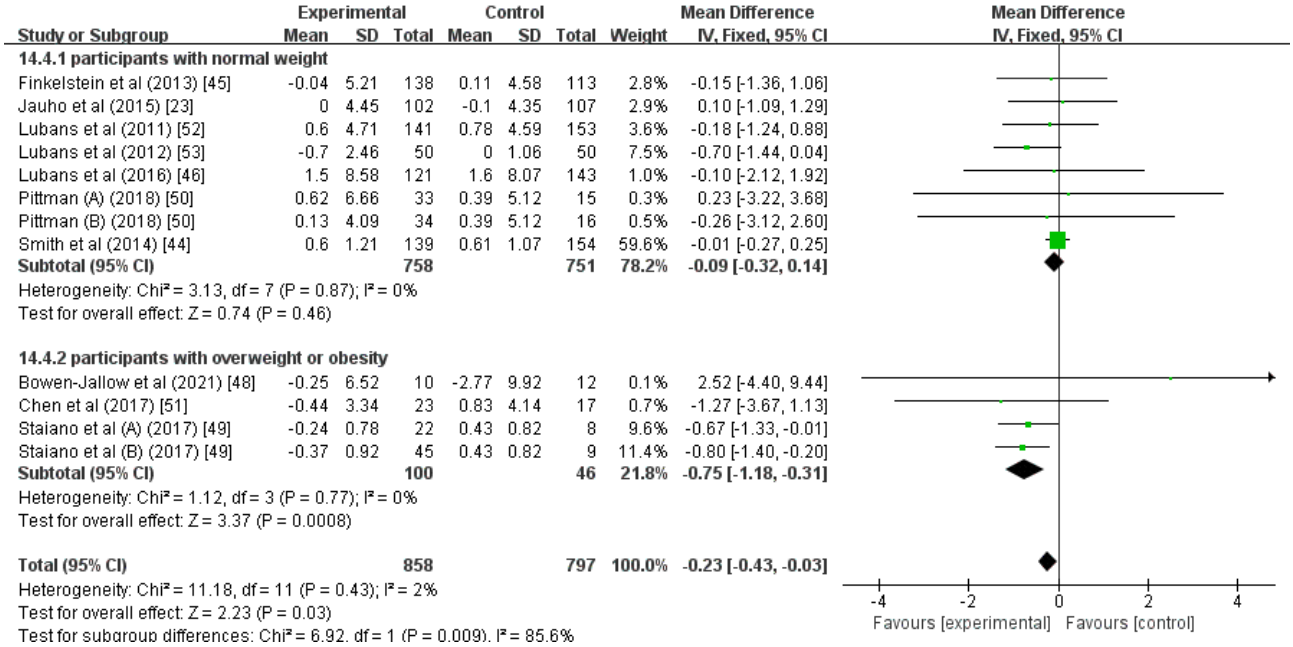


Figure 4. Forest plot of the intervention duration subgroup analysis.

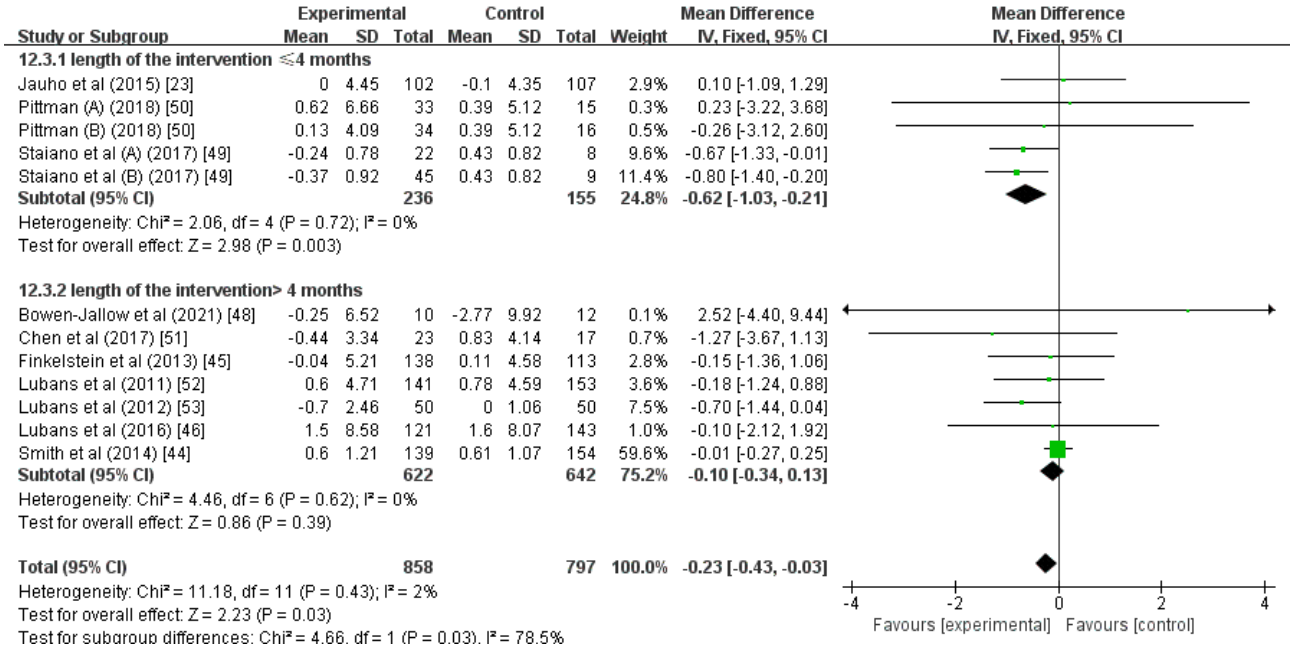
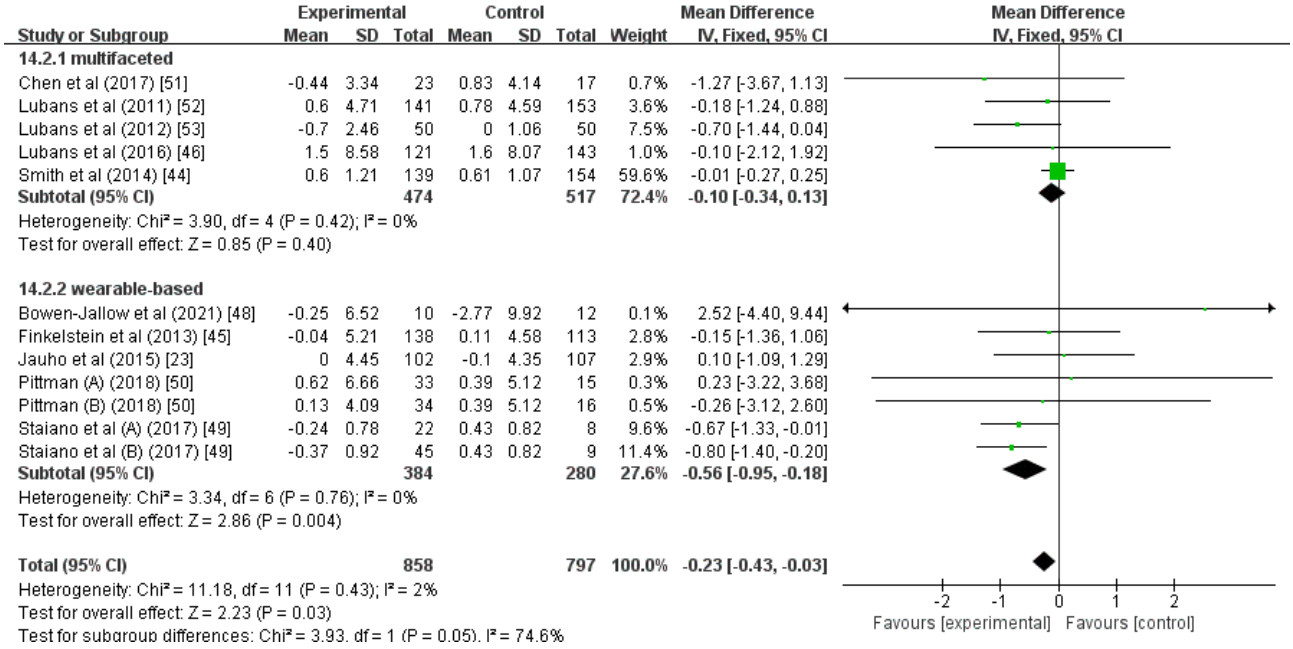


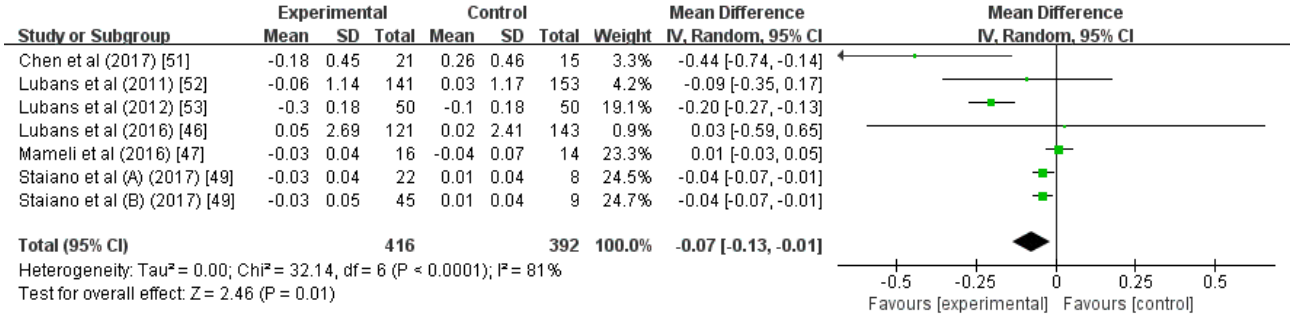
Figure 5. Forest plot of the intervention program subgroup analysis.



BMI-Z

A total of 6 studies explored the effects of wearable devices as physical activity interventions on the BMI-Z of children and adolescents [46,47,49,51-53]. There were statistically significant decreases in BMI-Z between the groups with wearable device interventions and the control group (MD –0.07; 95% CI –0.13 to –0.01; *P*=.01); heterogeneity was high and significant (I^2=81%; *P*<.01; Figure 6). Further sensitivity analysis was carried out. When the study of Lubans et al [53] was excluded, the effects of the intervention study on the BMI-Z showed a significant change (MD –0.03; 95% CI –0.07 to 0.01; *P*=.10; I^2=57%). Here, the Staiano et al [49] study had the greatest proportion (49.2%).

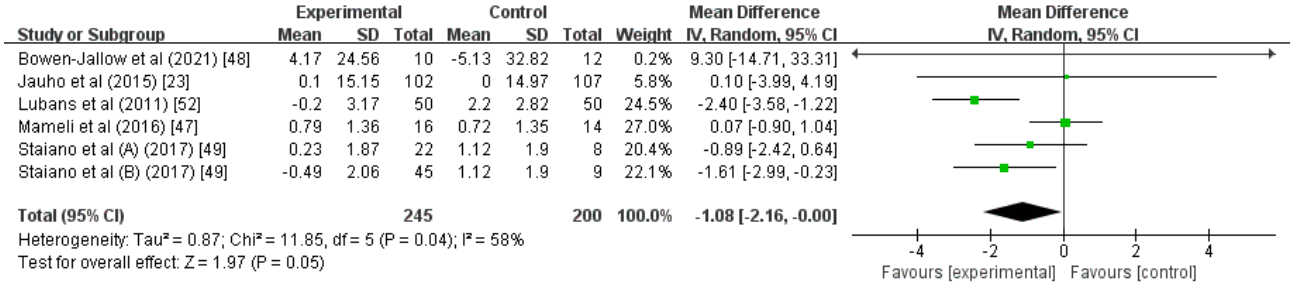
Figure 6. Forest plot of the effect of the wearable device interventions on BMI-Z (BMI z-score).



Body Weight

A total of 5 studies explored the effects of wearable devices as physical activity interventions on the body weight of children and adolescents [23,47-49,52]. Statistically significant decreases in body weight between the group with wearable device interventions and the control group were found (MD –1.08; 95% CI –2.16 to 0.00; *P*=.05); heterogeneity was high and significant (I^2=58%; *P*=.04; Figure 7). According to the results of further sensitivity analysis, after excluding the study performed by Mameli et al [47], the heterogeneity of the intervention study showed a significant change (MD –1.69; 95% CI –2.45 to –0.93; *P*<.01; I^2=0%). Among the studies, that by the Mameli et al [47] had the greatest proportion (27.0%).

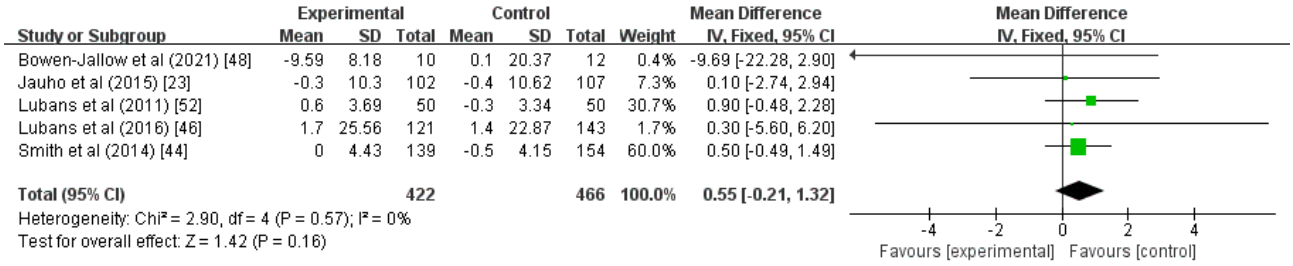
Figure 7. Forest plot of the effect of the wearable device interventions on body weight.



Waist Circumference

A total of 5 studies explored the effects of wearable devices as physical activity interventions on the waist circumference of children and adolescents [23,44,46,48,52]. Compared with the control group, the groups with wearable device interventions did not present statistically significant decreases in waist circumference (MD 0.55; 95% CI –0.21 to 1.32; *P*=.16); heterogeneity was low and insignificant (I^2=0%; *P*=.57; Figure 8). The study by Smith et al [44] had the greatest proportion in this analysis (60.0%).

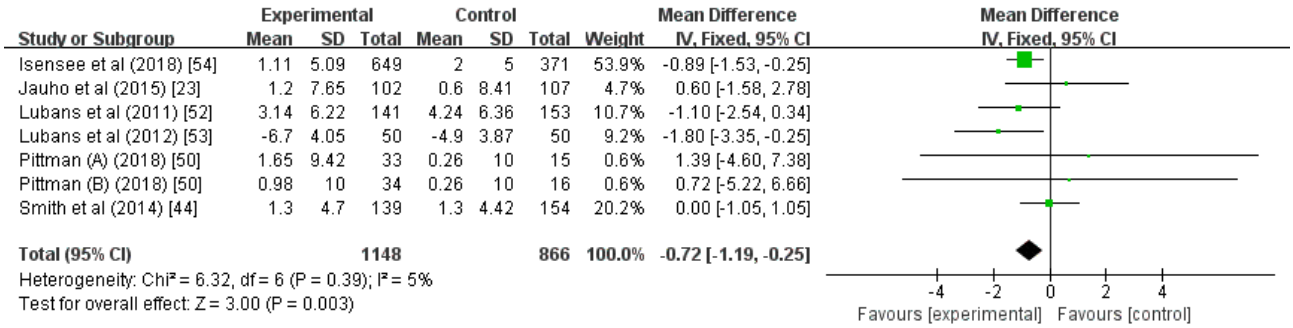
Figure 8. Forest plot of the effect of the wearable device interventions on waist circumference.



Body Fat Percentage

A total of 6 studies examined the effects of wearable devices as physical activity interventions on the body fat of children and adolescents [23,44,50,52-54]. There were statistically significant decreases in body fat between the group with wearable device interventions and the control group (MD –0.72; 95% CI –1.19 to –0.25; *P*=.003); heterogeneity was low and insignificant (I^2=5%; *P*=.39; Figure 9). The study by Isensee et al [54] had the greatest proportion (53.9%).

Figure 9. Forest plot of the effect of the wearable device interventions on body fat percentage.



Discussion

Principal Findings

This review and meta-analysis synthesized the existing evidence on the effectiveness of wearable devices as physical activity interventions on obesity-related outcomes in children and adolescents. The results indicated that, compared with the control group, wearable device interventions have statistically significant effects on BMI, BMI-Z, body weight, and body fat. However, no statistically significant effects on waist circumference were found. The subgroup analysis showed that for participants with overweight or obesity, in the short term, wearable-based interventions had a significantly greater intervention effect size on BMI.

This review demonstrates that the use of wearable devices as physical activity interventions can statistically significantly improve the BMI, BMI-Z, body weight, and body fat of children and adolescents. This is consistent with the results of other systematic reviews. Two previous reviews that examined the effectiveness of wearable devices as physical activity interventions for adults found that after using wearable devices, there were statistically significant improvements in obesity-related outcomes such as BMI and body weight [17,25]. These conclusions further confirmed the effect of physical activity on preventing and treating obesity. These conclusions

also demonstrate that physical activity plays an important role in the effectiveness of wearable devices for preventing and treating obesity. According to the behavior intervention technology model, wearable devices can promote physical activity through self-monitoring, goal setting, feedback, and motivation enhancement [55]. Assuming no drastic changes in dietary behavior, increasing physical activity, especially of moderate and vigorous intensity, will contribute to a negative caloric balance, leading to weight loss and BMI reduction [13,15,25]. Future research is needed to further explore what are the most effective physical activity indicators (eg, step count, total physical activity, and moderate-to-vigorous intensity physical activity) of wearable devices to prevent and treat obesity.

Four previous reviews examined the effectiveness of mobile health technology interventions in preventing and treating obesity in children and adolescents [56-59], finding that mobile health technology interventions yielded no significant improvements in obesity-related anthropometric outcomes such as BMI and body weight. These 4 reviews mixed different intervention strategies (promoting physical activity, monitoring food consumption, and strengthening communication and guidance) to explore the effects of mobile health technologies on prevention and treatment in obesity. By contrast, our review only concentrated on a single intervention strategy (promoting physical activity). These results may aid in the formulation of the most suitable intervention strategies on the use of mobile health technologies for addressing obesity among children and adolescents.

Waist circumference is a common indicator of abdominal obesity. This meta-analysis indicated that the wearable devices as physical activity interventions had no significant effect on waist circumference. This result is different to that of another meta-analysis in adults [17]. The reason may be that children and adolescents are in a sensitive period of growth and development, and waist circumference tends to increase, which will counteract part of the intervention effect [60]. It is worth noting that neither of these 2 systematic reviews included more than 5 RCTs. Thus, these conclusions must be treated cautiously.

A subgroup analysis found that, compared with participants with normal weight, those who were overweight or obese had a significantly greater intervention effect size on BMI. This result is consistent with other meta-analyses [17,25,26]. The reason may be that, compared with people with normal weight, the lifestyle of people with obesity is normally accompanied by more sedentary behaviors and less physical activity, and the level of physical activity is lower than that recommended by the World Health Organization [61,62]. Wearable devices can quantify the gap between the current activity level and the recommended amount [12,34]. Such a quantitative gap can motivate individuals to improve their physical activity level, and users can utilize self-regulation to change their physical activity habits as well as sedentary behaviors to manage their weight [13,63-67]. Wearable devices have achieved weight controlling effects by prompting a change in the lifestyle of people with obesity. For people of normal weight, the quantified physical activity level is not much different from the recommended amount. It cannot stimulate awareness of any deficiencies in their own activity, so the intervention effect is not obvious [68]. These facts suggest that the same intervention program may have different effects on different populations. Therefore, we need to analyze the baseline physical activity levels of different populations and formulate targeted intervention programs to achieve better intervention effects.

Another subgroup analysis found that interventions with a duration of ≤4 months had a significantly greater effect on BMI than those with a duration of >4 months. This is supported by the other systematic literature reviews [57,69], which showed that it is difficult for people to maintain focus on technology over time. Beyond 4 months, the freshness and interest in wearable devices disappear, resulting in the gradual disappearance of the intervention's effect [70]. However, 2 meta-analyses previously reported that wearable device interventions with the duration exceeding 4 months can achieve better BMI reduction [17,66]. It was preliminarily evidenced that children and adolescents were prone to losing interest and had poor compliance [71]. When using wearable devices for long-term interventions, targeted strategies should be applied at different periods of the intervention (such as self-monitoring in the first 3 months, peer competition at 3-6 months, family incentives at 6-9 months, and cash rewards at 9-12 months) so that children and adolescents can effectively maintain long-term interest and compliance.

The subgroup analysis clarified that, compared with the multifaceted intervention program, the wearable-based intervention program had a significantly greater impact on BMI. Wearable-based interventions focused on improving the user's physical activity levels to achieve weight loss. However, multifaceted interventions are based on multiple components. On the one hand, the intervention strategies did not focus on improving the physical activity level through wearable devices. On the other hand, the primary goal of intervention was not to prevent or treat obesity [37,57]. Therefore, the multifaceted intervention had a limited ability to reduce BMI. This suggests the need to focus on improving physical activity levels through wearable devices to prevent and treat obesity in children and adolescents.

Strengths and Limitations

Our systematic review has some strengths. First, to the best of our knowledge, this study may be the first meta-analysis to summarize the evidence on the effects of wearable devices as physical activity interventions on preventing and treating obesity in children and adolescents. Second, we chose the intervention tool focused on wearable devices, which are the latest mobile health technology products with advantages in functionality and convenience. Third, our review concentrated on a single intervention strategy (promoting physical activity through wearable devices). Fourth, the included studies were all RCTs with high-level evidence. Finally, this systematic review performed subgroup analyses to determine whether the characteristics of the interventions had an impact on the effect size.

The limitations of our review results must be clarified. First, relatively few studies met our inclusion criteria. This made it difficult to draw any definite conclusions. Second, 4 studies

were found to have a high risk of bias of incomplete outcome data, for which the dropout rates were >25%. Third, the high heterogeneity of the BMI-Z and body weight indicators in this meta-analysis cannot be ignored. Finally, in the meta-analysis of obesity-related anthropometric indicators, individual studies occupied excessive proportion in the analysis. Accordingly, the results of the meta-analysis may be affected by a single study.

Conclusions

This meta-analysis indicated that the use of wearable devices as physical activity interventions can significantly reduce BMI, BMI-Z, body weight, and body fat in children and adolescents, but failed to significantly improve waist circumference. The subgroup analysis showed that for participants with overweight or obesity, in short term, wearable-based interventions generally resulted in greater improvements in BMI. Therefore, wearable devices as physical activity interventions may be useful for preventing and treating obesity in children and adolescents. Future research is needed to identify the most effective physical activity indicators of wearable devices to prevent and treat obesity in children and adolescents.

Acknowledgments

This research was supported by the general research fund of the Zhejiang Provincial Department of Education, China (Y202045037).

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analysis) checklist.
[DOC File , 67 KB - mhealth_v10i4e32435_app1.doc]

Multimedia Appendix 2
PubMed search strategy.
[DOCX File , 15 KB - mhealth_v10i4e32435_app2.docx]

Multimedia Appendix 3
Risk of bias summary.
[PNG File , 12 KB - mhealth_v10i4e32435_app3.png]

Multimedia Appendix 4
Characteristics of included studies.
[DOCX File , 21 KB - mhealth_v10i4e32435_app4.docx]

References

1. Obesity and overweight. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2021-02-01]
2. Martens L, De Smet S, Yusof MYPM, Rajasekharan S. Association between overweight/obesity and periodontal disease in children and adolescents: a systematic review and meta-analysis. Eur Arch Paediatr Dent 2017 Apr;18(2):69-82. [doi: 10.1007/s40368-017-0272-1] [Medline: 28238103]
3. Reilly JJ, Kelly J. Long-term impact of overweight and obesity in childhood and adolescence on morbidity and premature mortality in adulthood: systematic review. Int J Obes (Lond) 2011 Jul;35(7):891-898. [doi: 10.1038/ijo.2010.222] [Medline: 20975725]
4. Twig G, Zucker I, Afek A, Cukierman-Yaffe T, Bendor CD, Derazne E, et al. Adolescent Obesity and Early-Onset Type 2 Diabetes. Diabetes Care 2020 Jul;43(7):1487-1495. [doi: 10.2337/dc19-1988] [Medline: 32321731]
5. Ye P, Chen F, Mi J. [Health hazards in childhood obesity: Evidence based on Chinese population]. Zhonghua Yu Fang Yi Xue Za Zhi 2016 Jan;50(1):97-100. [doi: 10.3760/cma.j.issn.0253-9624.2016.01.018] [Medline: 26792512]
6. Steinbeck KS. The importance of physical activity in the prevention of overweight and obesity in childhood: a review and an opinion. Obes Rev 2001 May;2(2):117-130. [doi: 10.1046/j.1467-789x.2001.00033.x] [Medline: 12119663]
7. Roberts SO. The role of physical activity in the prevention and treatment of childhood obesity. Pediatr Nurs 2000;26(1):33-6, 39. [Medline: 12026314]
8. Videira-Silva A, Hetherington-Rauth M, Sardinha LB, Fonseca H. Combined high-intensity interval training as an obesity-management strategy for adolescents. Eur J Sport Sci 2021 Nov 12:1-12. [doi: 10.1080/17461391.2021.1995508] [Medline: 34663193]

9. Martin A, Booth JN, Laird Y, Sproule J, Reilly JJ, Saunders DH. Physical activity, diet and other behavioural interventions for improving cognition and school achievement in children and adolescents with obesity or overweight. Cochrane Database Syst Rev 2018 Jan 29;1:CD009728 [FREE Full text] [doi: 10.1002/14651858.CD009728.pub3] [Medline: 29376563]
10. Videira-Silva A, Hetherington-Rauth M, Sardinha LB, Fonseca H. The effect of a physical activity consultation in the management of adolescent excess weight: Results from a non-randomized controlled trial. Clin Obes 2021 Dec;11(6):e12484. [doi: 10.1111/cob.12484] [Medline: 34467631]
11. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12:159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
12. Hao Y, Ma X, Zhu Z, Cao Z. Validity of Wrist-Wearable Activity Devices for Estimating Physical Activity in Adolescents: Comparative Study. JMIR Mhealth Uhealth 2021 Jan 07;9(1):e18320 [FREE Full text] [doi: 10.2196/18320] [Medline: 33410757]
13. Donnachie C, Wyke S, Mutrie N, Hunt K. 'It's like a personal motivator that you carried around wi' you': utilising self-determination theory to understand men's experiences of using pedometers to increase physical activity in a weight management programme. Int J Behav Nutr Phys Act 2017 May 05;14(1):61 [FREE Full text] [doi: 10.1186/s12966-017-0505-z] [Medline: 28476118]
14. Hu R, van Velthoven MH, Meinert E. Perspectives of People Who Are Overweight and Obese on Using Wearable Technology for Weight Management: Systematic Review. JMIR Mhealth Uhealth 2020 Jan 13;8(1):e12651 [FREE Full text] [doi: 10.2196/12651] [Medline: 31929104]
15. Lyons EJ, Swartz MC, Lewis ZH, Martinez E, Jennings K. Feasibility and Acceptability of a Wearable Technology Physical Activity Intervention With Telephone Counseling for Mid-Aged and Older Adults: A Randomized Controlled Pilot Trial. JMIR Mhealth Uhealth 2017 Mar 06;5(3):e28 [FREE Full text] [doi: 10.2196/mhealth.6967] [Medline: 28264796]
16. Allen JK, Stephens J, Dennison Himmelfarb CR, Stewart KJ, Hauck S. Randomized controlled pilot study testing use of smartphone technology for obesity treatment. J Obes 2013;2013:151597 [FREE Full text] [doi: 10.1155/2013/151597] [Medline: 24392223]
17. Yen H, Chiu H. The effectiveness of wearable technologies as physical activity interventions in weight control: A systematic review and meta-analysis of randomized controlled trials. Obes Rev 2019 Oct;20(10):1485-1493. [doi: 10.1111/obr.12909] [Medline: 31342646]
18. Puigdomenech Puig E, Robles N, Saigí-Rubió F, Zamora A, Moharra M, Paluzie G, et al. Assessment of the Efficacy, Safety, and Effectiveness of Weight Control and Obesity Management Mobile Health Interventions: Systematic Review. JMIR Mhealth Uhealth 2019 Oct 25;7(10):e12612 [FREE Full text] [doi: 10.2196/12612] [Medline: 31654566]
19. Lugones-Sanchez C, Sanchez-Calavera MA, Repiso-Gento I, Adalia EG, Ramirez-Manent JI, Agudo-Conde C, EVIDENT 3 Investigators. Effectiveness of an mHealth Intervention Combining a Smartphone App and Smart Band on Body Composition in an Overweight and Obese Population: Randomized Controlled Trial (EVIDENT 3 Study). JMIR Mhealth Uhealth 2020 Nov 26;8(11):e21771 [FREE Full text] [doi: 10.2196/21771] [Medline: 33242020]
20. Kim J, Ryu B, Cho S, Heo E, Kim Y, Lee J, et al. Impact of Personal Health Records and Wearables on Health Outcomes and Patient Response: Three-Arm Randomized Controlled Trial. JMIR Mhealth Uhealth 2019 Jan 04;7(1):e12070 [FREE Full text] [doi: 10.2196/12070] [Medline: 30609978]
21. Lee J, Kang M, Lee S. Effects of the e-Motivate4Change Program on Metabolic Syndrome in Young Adults Using Health Apps and Wearable Devices: Quasi-Experimental Study. J Med Internet Res 2020 Jul 30;22(7):e17031 [FREE Full text] [doi: 10.2196/17031] [Medline: 32729838]
22. Butte KD, Bahmani A, Butte AJ, Li X, Snyder MP. Five-year pediatric use of a digital wearable fitness device: lessons from a pilot case study. JAMIA Open 2021 Jul;4(3):ooab054 [FREE Full text] [doi: 10.1093/jamiaopen/ooab054] [Medline: 34350390]
23. Jauho A, Pyky R, Ahola R, Kangas M, Virtanen P, Korpelainen R, et al. Effect of wrist-worn activity monitor feedback on physical activity behavior: A randomized controlled trial in Finnish young men. Prev Med Rep 2015;2:628-634 [FREE Full text] [doi: 10.1016/j.pmedr.2015.07.005] [Medline: 26844128]
24. Davergne T, Pallot A, Dechartres A, Fautrel B, Gossec L. Use of Wearable Activity Trackers to Improve Physical Activity Behavior in Patients With Rheumatic and Musculoskeletal Diseases: A Systematic Review and Meta-Analysis. Arthritis Care Res (Hoboken) 2019 Jun;71(6):758-767. [doi: 10.1002/acr.23752] [Medline: 30221489]
25. McDonough DJ, Su X, Gao Z. Health wearable devices for weight and BMI reduction in individuals with overweight/obesity and chronic comorbidities: systematic review and network meta-analysis. Br J Sports Med 2021 Aug;55(16):917-925. [doi: 10.1136/bjsports-2020-103594] [Medline: 33731385]
26. Tang MSS, Moore K, McGavigan A, Clark RA, Ganesan AN. Effectiveness of Wearable Trackers on Physical Activity in Healthy Adults: Systematic Review and Meta-Analysis of Randomized Controlled Trials. JMIR Mhealth Uhealth 2020 Jul 22;8(7):e15576 [FREE Full text] [doi: 10.2196/15576] [Medline: 32706685]
27. Mackintosh KA, Chappel SE, Salmon J, Timperio A, Ball K, Brown H, et al. Parental Perspectives of a Wearable Activity Tracker for Children Younger Than 13 Years: Acceptability and Usability Study. JMIR Mhealth Uhealth 2019 Nov 04;7(11):e13858 [FREE Full text] [doi: 10.2196/13858] [Medline: 31682585]

28. Ridgers ND, Timperio A, Brown H, Ball K, Macfarlane S, Lai SK, et al. Wearable Activity Tracker Use Among Australian Adolescents: Usability and Acceptability Study. JMIR Mhealth Uhealth 2018 Apr 11;6(4):e86 [FREE Full text] [doi: 10.2196/mhealth.9199] [Medline: 29643054]
29. Oliveira RGD, Guedes DP. Physical Activity, Sedentary Behavior, Cardiorespiratory Fitness and Metabolic Syndrome in Adolescents: Systematic Review and Meta-Analysis of Observational Evidence. PLoS One 2016;11(12):e0168503 [FREE Full text] [doi: 10.1371/journal.pone.0168503] [Medline: 27997601]
30. Jurado-Castro JM, Gil-Campos M, Gonzalez-Gonzalez H, Llorente-Cantarero FJ. Evaluation of Physical Activity and Lifestyle Interventions Focused on School Children with Obesity Using Accelerometry: A Systematic Review and Meta-Analysis. Int J Environ Res Public Health 2020 Aug 19;17(17):6031-6046 [FREE Full text] [doi: 10.3390/ijerph17176031] [Medline: 32825085]
31. El-Amrawy F, Nounou MI. Are Currently Available Wearable Devices for Activity Tracking and Heart Rate Monitoring Accurate, Precise, and Medically Beneficial? Healthc Inform Res 2015 Oct;21(4):315-320 [FREE Full text] [doi: 10.4258/hir.2015.21.4.315] [Medline: 26618039]
32. Brooke SM, An H, Kang S, Noble JM, Berg KE, Lee J. Concurrent Validity of Wearable Activity Trackers Under Free-Living Conditions. J Strength Cond Res 2017 Apr;31(4):1097-1106. [doi: 10.1519/JSC.0000000000001571] [Medline: 27465631]
33. Ferguson T, Rowlands AV, Olds T, Maher C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. Int J Behav Nutr Phys Act 2015 Mar 27;12:42 [FREE Full text] [doi: 10.1186/s12966-015-0201-9] [Medline: 25890168]
34. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, et al. Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate: Systematic Review. JMIR Mhealth Uhealth 2020 Sep 08;8(9):e18694 [FREE Full text] [doi: 10.2196/18694] [Medline: 32897239]
35. Dorn D, Gorzelitz J, Gangnon R, Bell D, Koltyn K, Cadmus-Bertram L. Automatic Identification of Physical Activity Type and Duration by Wearable Activity Trackers: A Validation Study. JMIR Mhealth Uhealth 2019 May 23;7(5):e13547 [FREE Full text] [doi: 10.2196/13547] [Medline: 31124470]
36. Mercer K, Li M, Giangregorio L, Burns C, Grindrod K. Behavior Change Techniques Present in Wearable Activity Trackers: A Critical Analysis. JMIR Mhealth Uhealth 2016 Apr 27;4(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4461] [Medline: 27122452]
37. Brickwood K, Watson G, O'Brien J, Williams AD. Consumer-Based Wearable Activity Trackers Increase Physical Activity Participation: Systematic Review and Meta-Analysis. JMIR Mhealth Uhealth 2019 Apr 12;7(4):e11819 [FREE Full text] [doi: 10.2196/11819] [Medline: 30977740]
38. Ringeval M, Wagner G, Denford J, Paré G, Kitsiou S. Fitbit-Based Interventions for Healthy Lifestyle Outcomes: Systematic Review and Meta-Analysis. J Med Internet Res 2020 Oct 12;22(10):e23954 [FREE Full text] [doi: 10.2196/23954] [Medline: 33044175]
39. Eckerstorfer LV, Tanzer NK, Vogrincic-Haselbacher C, Kedia G, Brohmer H, Dinslaken I, et al. Key Elements of mHealth Interventions to Successfully Increase Physical Activity: Meta-Regression. JMIR Mhealth Uhealth 2018 Nov 12;6(11):e10076 [FREE Full text] [doi: 10.2196/10076] [Medline: 30425028]
40. Düking P, Tafler M, Wallmann-Sperlich B, Sperlich B, Kleih S. Behavior Change Techniques in Wrist-Worn Wearables to Promote Physical Activity: Content Analysis. JMIR Mhealth Uhealth 2020 Nov 19;8(11):e20820 [FREE Full text] [doi: 10.2196/20820] [Medline: 33211023]
41. Creaser AV, Clemes SA, Costa S, Hall J, Ridgers ND, Barber SE, et al. The Acceptability, Feasibility, and Effectiveness of Wearable Activity Trackers for Increasing Physical Activity in Children and Adolescents: A Systematic Review. Int J Environ Res Public Health 2021 Jun 08;18(12):6211-6239 [FREE Full text] [doi: 10.3390/ijerph18126211] [Medline: 34201248]
42. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
43. Cumpston M, Li T, Page MJ, Chandler J, Welch VA, Higgins JP, et al. Updated guidance for trusted systematic reviews: a new edition of the Cochrane Handbook for Systematic Reviews of Interventions. Cochrane Database Syst Rev 2019 Oct 03;10:ED000142. [doi: 10.1002/14651858.ED000142] [Medline: 31643080]
44. Smith JJ, Morgan PJ, Plotnikoff RC, Dally KA, Salmon J, Okely AD, et al. Smart-phone obesity prevention trial for adolescent boys in low-income communities: the ATLAS RCT. Pediatrics 2014 Sep;134(3):e723-e731. [doi: 10.1542/peds.2014-1012] [Medline: 25157000]
45. Finkelstein EA, Tan Y, Malhotra R, Lee C, Goh S, Saw S. A cluster randomized controlled trial of an incentive-based outdoor physical activity program. J Pediatr 2013 Jul;163(1):167-72.e1. [doi: 10.1016/j.jpeds.2013.01.009] [Medline: 23415616]
46. Lubans DR, Smith JJ, Plotnikoff RC, Dally KA, Okely AD, Salmon J, et al. Assessing the sustained impact of a school-based obesity prevention program for adolescent boys: the ATLAS cluster randomized controlled trial. Int J Behav Nutr Phys Act 2016 Aug 20;13:92 [FREE Full text] [doi: 10.1186/s12966-016-0420-8] [Medline: 27542825]

47. Mameli C, Brunetti D, Colombo V, Bedogni G, Schneider L, Penagini F, et al. Combined use of a wristband and a smartphone to reduce body weight in obese children: randomized controlled trial. Pediatr Obes 2018 Feb;13(2):81-87. [doi: 10.1111/ijpo.12201] [Medline: 27900849]
48. Bowen-Jallow K, Nunez-Lopez O, Wright A, Fuchs E, Ahn M, Lyons E, et al. Wearable Activity Tracking Device Use in an Adolescent Weight Management Clinic: A Randomized Controlled Pilot Trial. J Obes 2021;2021:7625034 [FREE Full text] [doi: 10.1155/2021/7625034] [Medline: 33505717]
49. Staiano AE, Beyl RA, Hsia DS, Jarrell AR, Katzmarzyk PT, Mantzor S, et al. Step Tracking with Goals Increases Children's Weight Loss in Behavioral Intervention. Child Obes 2017 Aug;13(4):283-290 [FREE Full text] [doi: 10.1089/chi.2017.0047] [Medline: 28440662]
50. Pittman AF. Effect of a School-Based Activity Tracker, Companion Social Website, and Text Messaging Intervention on Exercise, Fitness, and Physical Activity Self-Efficacy of Middle School Students. J Sch Nurs 2020 Apr;36(2):112-120. [doi: 10.1177/1059840518791223] [Medline: 30068252]
51. Chen J, Guedes CM, Cooper BA, Lung AE. Short-Term Efficacy of an Innovative Mobile Phone Technology-Based Intervention for Weight Management for Overweight and Obese Adolescents: Pilot Study. Interact J Med Res 2017 Aug 02;6(2):e12 [FREE Full text] [doi: 10.2196/ijmr.7860] [Medline: 28768612]
52. Lubans DR, Morgan PJ, Aguiar EJ, Callister R. Randomized controlled trial of the Physical Activity Leaders (PALs) program for adolescent boys from disadvantaged secondary schools. Prev Med 2011;52(3-4):239-246. [doi: 10.1016/j.ypmed.2011.01.009] [Medline: 21276812]
53. Lubans DR, Morgan PJ, Okely AD, Dewar D, Collins CE, Batterham M, et al. Preventing Obesity Among Adolescent Girls: One-Year Outcomes of the Nutrition and Enjoyable Activity for Teen Girls (NEAT Girls) Cluster Randomized Controlled Trial. Arch Pediatr Adolesc Med 2012 Sep 01;166(9):821-827. [doi: 10.1001/archpediatrics.2012.41] [Medline: 22566517]
54. Isensee B, Suchert V, Hansen J, Weisser B, Hanewinkel R. Effects of a School-Based Pedometer Intervention in Adolescents: 1-Year Follow-Up of a Cluster-Randomized Controlled Trial. J Sch Health 2018 Oct;88(10):717-724. [doi: 10.1111/josh.12676] [Medline: 30203479]
55. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 05;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
56. Quelly SB, Norris AE, DiPietro JL. Impact of mobile apps to combat obesity in children and adolescents: A systematic literature review. J Spec Pediatr Nurs 2016 Jan;21(1):5-17. [doi: 10.1111/jspn.12134] [Medline: 26494019]
57. Shin Y, Kim SK, Lee M. Mobile phone interventions to improve adolescents' physical health: A systematic review and meta-analysis. Public Health Nurs 2019 Nov;36(6):787-799. [doi: 10.1111/phn.12655] [Medline: 31456259]
58. Turner T, Spruijt-Metz D, Wen CKF, Hingle MD. Prevention and treatment of pediatric obesity using mobile and wireless technologies: a systematic review. Pediatr Obes 2015 Dec;10(6):403-409 [FREE Full text] [doi: 10.1111/ijpo.12002] [Medline: 25641770]
59. Tully L, Burls A, Sorensen J, El-Moslemany R, O'Malley G. Mobile Health for Pediatric Weight Management: Systematic Scoping Review. JMIR Mhealth Uhealth 2020 Jun 03;8(6):e16214 [FREE Full text] [doi: 10.2196/16214] [Medline: 32490849]
60. Videira-Silva A, Fonseca H. The effect of a physical activity consultation on body mass index z-score of overweight adolescents: results from a pediatric outpatient obesity clinic. Eur J Pediatr 2017 May;176(5):655-660. [doi: 10.1007/s00431-017-2892-1] [Medline: 28275861]
61. Kirk MA, Amiri M, Pirbaglou M, Ritvo P. Wearable Technology and Physical Activity Behavior Change in Adults With Chronic Cardiometabolic Disease: A Systematic Review and Meta-Analysis. Am J Health Promot 2019 Jun;33(5):778-791. [doi: 10.1177/0890117118816278] [Medline: 30586996]
62. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008 Jan;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3] [Medline: 18091006]
63. Slootmaker SM, Chinapaw MJM, Seidell JC, van Mechelen W, Schuit AJ. Accelerometers and Internet for physical activity promotion in youth? Feasibility and effectiveness of a minimal intervention [ISRCTN93896459]. Prev Med 2010 Jul;51(1):31-36. [doi: 10.1016/j.ypmed.2010.03.015] [Medline: 20380847]
64. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
65. Direito A, Dale LP, Shields E, Dobson R, Whittaker R, Maddison R. Do physical activity and dietary smartphone applications incorporate evidence-based behaviour change techniques? BMC Public Health 2014 Jun 25;14:646 [FREE Full text] [doi: 10.1186/1471-2458-14-646] [Medline: 24965805]
66. Cavero-Redondo I, Martinez-Vizcaino V, Fernandez-Rodriguez R, Saz-Lara A, Pascual-Morena C, Álvarez-Bueno C. Effect of Behavioral Weight Management Interventions Using Lifestyle mHealth Self-Monitoring on Weight Loss: A Systematic Review and Meta-Analysis. Nutrients 2020 Jul 03;12(7):1977-1996 [FREE Full text] [doi: 10.3390/nu12071977] [Medline: 32635174]

67. Ho TJH, Lee CCS, Wong SN, Lau Y. Internet-based self-monitoring interventions for overweight and obese adolescents: A systematic review and meta-analysis. Int J Med Inform 2018 Dec;120:20-30. [doi: 10.1016/j.ijmedinf.2018.09.019] [Medline: 30409343]
68. Metcalf B, Henley W, Wilkin T. Effectiveness of intervention on physical activity of children: systematic review and meta-analysis of controlled trials with objectively measured outcomes (EarlyBird 54). BMJ 2012 Sep 27;345:e5888. [doi: 10.1136/bmj.e5888] [Medline: 23044984]
69. Ridgers ND, McNarry MA, Mackintosh KA. Feasibility and Effectiveness of Using Wearable Activity Trackers in Youth: A Systematic Review. JMIR Mhealth Uhealth 2016 Nov 23;4(4):e129 [FREE Full text] [doi: 10.2196/mhealth.6540] [Medline: 27881359]
70. Chong KPL, Guo JZ, Deng X, Woo BKP. Consumer Perceptions of Wearable Technology Devices: Retrospective Review and Analysis. JMIR Mhealth Uhealth 2020 Apr 20;8(4):e17544 [FREE Full text] [doi: 10.2196/17544] [Medline: 32310148]
71. Schaefer SE, Ching CC, Breen H, German JB. Wearing, Thinking, and Moving: Testing the Feasibility of Fitness Tracking with Urban Youth. American Journal of Health Education 2016 Jan 08;47(1):8-16. [doi: 10.1080/19325037.2015.1111174]

Abbreviations

PRISMA: Preferred Reporting Items for Systematic Review and Meta-Analysis
RCT: randomized controlled trial

Edited by L Buis, A Mavragani; submitted 28.07.21; peer-reviewed by V Martinez-Vizcaino, A Videira-Silva; comments to author 28.10.21; revised version received 15.12.21; accepted 21.03.22; published 08.04.22.

Please cite as:
Wang W, Cheng J, Song W, Shen Y
The Effectiveness of Wearable Devices as Physical Activity Interventions for Preventing and Treating Obesity in Children and Adolescents: Systematic Review and Meta-analysis
JMIR Mhealth Uhealth 2022;10(4):e32435
URL: https://mhealth.jmir.org/2022/4/e32435
doi:10.2196/32435
PMID:35394447

Review

Loneliness and Social Isolation Detection Using Passive Sensing Techniques: Scoping Review

Malik Muhammad Qirtas[1], BE, MS; Evi Zafeiridi[1], PhD; Dirk Pesch[1], PhD; Eleanor Bantry White[1], PhD

School of Computer Science & Information Technology, University College Cork, Cork, Ireland

Corresponding Author:
Malik Muhammad Qirtas, BE, MS
School of Computer Science & Information Technology
University College Cork
Room 2.09
Western Gateway Building, Western Road
Cork, T23 W623
Ireland
Phone: 353 0851873544
Email: qirtas333@gmail.com

Abstract

Background: Loneliness and social isolation are associated with multiple health problems, including depression, functional impairment, and death. Mobile sensing using smartphones and wearable devices, such as fitness trackers or smartwatches, as well as ambient sensors, can be used to acquire data remotely on individuals and their daily routines and behaviors in real time. This has opened new possibilities for the early detection of health and social problems, including loneliness and social isolation.

Objective: This scoping review aimed to identify and synthesize recent scientific studies that used passive sensing techniques, such as the use of in-home ambient sensors, smartphones, and wearable device sensors, to collect data on device users' daily routines and behaviors to detect loneliness or social isolation. This review also aimed to examine various aspects of these studies, especially target populations, privacy, and validation issues.

Methods: A scoping review was undertaken, following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews). Studies on the topic under investigation were identified through 6 databases (IEEE Xplore, Scopus, ACM, PubMed, Web of Science, and Embase). The identified studies were screened for the type of passive sensing detection methods for loneliness and social isolation, targeted population, reliability of the detection systems, challenges, and limitations of these detection systems.

Results: After conducting the initial search, a total of 40,071 papers were identified. After screening for inclusion and exclusion criteria, 29 (0.07%) studies were included in this scoping review. Most studies (20/29, 69%) used smartphone and wearable technology to detect loneliness or social isolation, and 72% (21/29) of the studies used a validated reference standard to assess the accuracy of passively collected data for detecting loneliness or social isolation.

Conclusions: Despite the growing use of passive sensing technologies for detecting loneliness and social isolation, some substantial gaps still remain in this domain. A population heterogeneity issue exists among several studies, indicating that different demographic characteristics, such as age and differences in participants' behaviors, can affect loneliness and social isolation. In addition, despite extensive personal data collection, relatively few studies have addressed privacy and ethical issues. This review provides uncertain evidence regarding the use of passive sensing to detect loneliness and social isolation. Future research is needed using robust study designs, measures, and examinations of privacy and ethical concerns.

(JMIR Mhealth Uhealth 2022;10(4):e34638) doi:10.2196/34638

KEYWORDS

passive sensing; loneliness; social isolation; smartphone; sensors; wearables; monitoring; scoping review; eHealth; mHealth; mobile phone

Introduction

Background

Humans are social creatures; thus, maintaining a healthy and positive social interaction is a necessary component of human existence [1]. Over the past few decades, loneliness has emerged as a global issue, and people rate it as a primary source of unhappiness in their lives [2]. Perlman defined loneliness as when an individual thinks that the quality or quantity of their social relationships is inadequate [3]. Loneliness or social isolation is a common problem that most people experience at some stage in their lives; however, it can be very distressing, especially when it becomes chronic [4]. In other words, loneliness is a distressing emotional state in which a person is dissatisfied with the proximity and pattern of their social interactions and relationships. Thus, loneliness is a fully subjective state of mind; a person can live completely alone and not experience loneliness, whereas another person may experience loneliness despite having a wide social network. Adult loneliness is a growing concern; according to a survey from the United Kingdom, 1 in every 20 adults reports feeling completely alone [5].

Social isolation refers to a person's lack of social interaction with others at the individual or group level [6]. Social isolation can occur voluntarily or involuntarily and can have a positive or negative effect on an individual based on their mental and physical health [7]. Social isolation and loneliness are overlapping concepts often used interchangeably. Unlike social isolation, loneliness is always involuntary and involves negative emotions [8]. Although these two ideas are closely related, they do not have the same meaning. An individual can feel alone in a crowd, whereas another person can benefit from social isolation while experiencing solitude. Loneliness may occur when an individual is socially isolated for a prolonged period [9]. Loneliness is also a complex phenomenon that varies in intensity and is influenced by a variety of factors and conditions. To explain the multifaceted aspects of loneliness, Sadler and Weiss [10] distinguished between *emotional loneliness* and *social loneliness*. Emotional loneliness refers to the absence of close or intimate connections, whereas social loneliness refers to the absence of social networks. For instance, a child who has lost their mother experiences loneliness differently from a child who lacks playmates [11].

Loneliness is influenced by a variety of factors, including age, poor health, physical disability, relationship status, living alone, infertility, low wages, low levels of education, and socioeconomic status [12]. Certain personality traits, such as lower levels of extraversion and higher levels of neuroticism, may also increase the risk of loneliness [13]. According to a study on loneliness over the life span, it peaks in late adolescence, gradually declines during middle age, and then returns in later life [14]. As people age, they often lose social contact because of illness and cognitive decline, leading to loss of social relationships. Gradually, their social networks decline as they grow older, resulting in very few people from whom they receive support [15]. A report published by the Survey of Health, Ageing, and Retirement in Europe found that older adults living alone can experience higher levels of loneliness [16].

Loneliness has a detrimental effect on physical and mental health, increasing an individual's risk of morbidity and mortality [17]. The effects of loneliness on cardiovascular health have been widely studied [18]. Poor social bonds have been linked to a 29% increase in incident coronary heart disease and a 32% increase in stroke according to a meta-analysis of longitudinal data on 35,925 people [19]. Loneliness can also result in various neuroendocrine problems and a weakened immune system [20]. In addition, loneliness can disrupt normal blood pressure, sleep patterns, and cortisol levels, resulting in serious health problems or even death [21]. Loneliness has been linked to various detrimental mental health outcomes [22], including depression [23], suicidality, less positive mood, poor sleep quality, poor overall physical health, and physiological abnormalities [13].

In addition to their primary function as communication devices, modern smartphones provide a plethora of capabilities [24]. A smartphone can be used as a handheld camera, a navigator, a fitness tracker, and a personal assistant [24]. As smartphones are equipped with a variety of powerful sensors, they have evolved into pervasive tracking systems [25]. Similarly, sensors in new wearable devices, such as smart watches and fitness trackers, have created the possibility of transforming them into a robust health tracking system [26]. Among the many fields that make use of the several capabilities of smartphones and wearables, one is passive sensing, the process of collecting data in the background using the ubiquitous nature of smartphones and wearables without the user's active engagement. This concept of passive sensing or self-monitoring emerged from studies in the field of ubiquitous computing, where it has been called *context-aware computing* [27]. Smartphones can collect a vast amount of data regarding a user's behavioral patterns, which can be modeled into bioindicators of the user's well-being [28]. Using sensors, such as an accelerometer, a heart rate sensor, or a microphone as well as the GPS or Bluetooth connectivity as a proximity sensor, researchers collect high volumes of data about a user's everyday life and behavioral patterns, such as social activities, time spent at certain places, daily health data, and a user's activity log. All these capabilities of smartphones and wearables have made them a promising tool for tracking users' health and well-being on a ubiquitous and passive level. In addition to smartphones and wearables, various environmental or ambient sensor-based devices have been used to passively gather information about users' in-home behavioral habits, such as passive and active infrared sensors, pressure mats and tiles, and camera and microphone sensors, particularly for older adults [29].

In contrast to systematic reviews and other review-based approaches, scoping reviews are a relatively new tool for searching and summarizing the literature. Systematic reviews address specific questions using predefined methods to assess study quality. Scoping reviews can be used to map new or emerging fields of research and are useful when investigating a broader set of review questions that require the inclusion of studies from a wider range of study methodologies [30-32]. Researchers can conduct a scoping analysis in a recent or neglected research area for four primary reasons: to determine

the design and purpose of the study, to determine the necessity of conducting a systematic review, to summarize the research results, and to locate research gaps within the existing literature [33]. This study investigated the broader field of loneliness and social isolation detection through passive sensing as a recent and less explored research area. This review aims to explore the types of passive sensing detection methods available in the literature for loneliness and social isolation, to investigate the target populations of these detection methods, to review how the developed systems were validated, and to find challenges and limitations of these detection systems.

Loneliness Detection Approaches: An Overview

Assessing Loneliness in Research and Clinical Settings

There are several tools that assess loneliness. These assessments or scales are well established and have shown promising results in inferring loneliness. The University of California, Los Angeles (UCLA) loneliness measurement scale (version 3) is one of the most widely used in the general population and health care settings for assessing the frequency and severity of subjective loneliness in an individual [34-37]. Several other measures are available that may be used as a validated reference measure to determine an individual's level of loneliness [38-41]. These scales include questions regarding marital status, social involvement in religious activities and clubs, social contact, social support, and social networks [28,42-46]. Most loneliness measures appear reliable. However, the evidence for validity is relatively limited, consisting mostly of comparisons between *at-risk* and normal populations or correlations with questions generating explicit self-identification as being lonely [47]. Moreover, we could not find any study that has used passive sensing methods to detect loneliness or social isolation in a larger, nonclinical population.

Technology-Based Detection Through Passive Sensing

Over the last decade, there has been increasing interest among researchers and clinicians in the opportunities presented by technology-based approaches for detecting loneliness and social isolation. Innovative technology-based approaches are extremely effective in terms of ubiquity and passive sensing [48] because they do not require the participant's active participation. In the existing literature, most studies have used the following two methods to detect loneliness and social isolation: smart home–based methods (ambient sensors–based) and smartphone and wearable-based mobile sensing methods.

A smart home comprises a variety of in-home environmental or ambient sensors, along with more specialized audio, video, and biometric systems that can be used by family and caregivers to track older individuals' actions and well-being while being physically away. A smart home can include a variety of sensors depending on the application. For instance, numerous studies have used in-home surveillance through video cameras [49,50], whereas others have used body-worn tags [51,52]. On the other hand, inexpensive ambient sensors allow a more unobtrusive method of monitoring behaviors in the home without user involvement. Ambient in-home sensors have been used for human behavior learning over the last few years, in which emotions, daily life patterns, or personality could be related to loneliness levels [53-55]. It incorporates an assessment of both physical and emotional well-being. These ambient in-home sensors are cost-effective, energy efficient, and easy to mount and maintain. They include motion sensors, which emit a signal whenever a motion is detected within the sensor's coverage range; touch sensors, which generate a signal whenever a door is opened or closed; and pressure sensors, which emit a signal whenever a pressure threshold is crossed at the location of the sensor, which is usually in beds. Consequently, these sensors collect data on a user's various behaviors, such as time spent in various areas of the home, sleep habits, time spent inside and outside the home, and in-home mobility patterns. These activity patterns can ultimately serve as measures or biomarkers of loneliness or social isolation.

Researchers have also used mobile and wearable devices for health monitoring over the last few years. The rapid proliferation of smartphones and wearables, such as fitness trackers, which are equipped with powerful sensors, may provide another pathway for detecting loneliness and social isolation. The data passively acquired from mobile sensors can be modeled into various behavioral habits that can be used to identify lonely individuals. Behavioral patterns included a participant's social experiences, mobility patterns, and frequently visited places.

Methods

Search Strategy and Data Sources

The protocol for this scoping review was developed as per the guidelines of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [56] and PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [57]. The databases IEEE Xplore, Scopus, ACM, PubMed, Web of Science, and Embase were searched. Articles published between January 2011 and December 2021 were extracted independently by the authors MMQ and EZ. The search terms and results for each database are presented in Table 1. The search criteria were based on loneliness, social isolation, and detection-related keywords.

Table 1. Search keywords and result statistics for computer science and social science databases.

Database and keywords combination	Search results, n
IEEE Xplore	
Loneliness AND Detection	13
Loneliness AND Sensing	27
Loneliness AND Passive Sensing	1
Loneliness AND Monitoring	15
Social Isolation AND Detection	56
Social Isolation AND Sensing	41
Social Isolation AND Passive Sensing	1
Social Isolation AND Monitoring	65
Scopus	
Loneliness AND Detection	4373
Loneliness AND Sensing	725
Loneliness AND Passive Sensing	61
Loneliness AND Monitoring	6443
Social Isolation AND Detection	5501
Social Isolation AND Sensing	834
Social Isolation AND Passive Sensing	51
Social Isolation AND Monitoring	5901
ACM	
Loneliness AND Detection	703
Loneliness AND Sensing	1339
Loneliness AND Passive Sensing	27
Loneliness AND Monitoring	603
Social Isolation AND Detection	313
Social Isolation AND Sensing	631
Social Isolation AND Passive Sensing	10
Social Isolation AND Monitoring	357
PubMed	
Loneliness AND Detection	220
Loneliness AND Sensing	770
Loneliness AND Passive Sensing	8
Loneliness AND Monitoring	200
Social Isolation AND Detection	2238
Social Isolation AND Sensing	1981
Social Isolation AND Passive Sensing	24
Social Isolation AND Monitoring	1202
Web of Science	
Loneliness AND Detection	109
Loneliness AND Sensing	1130
Loneliness AND Passive Sensing	16
Loneliness AND Monitoring	299
Social Isolation AND Detection	350

Database and keywords combination	Search results, n
Social Isolation AND Sensing	1195
Social Isolation AND Passive Sensing	13
Social Isolation AND Monitoring	701
Embase	
Loneliness AND Detection	137
Loneliness AND Sensing	22
Loneliness AND Passive Sensing	1
Loneliness AND Monitoring	235
Social Isolation AND Detection	390
Social Isolation AND Sensing	34
Social Isolation AND Passive Sensing	1
Social Isolation AND Monitoring	704

Inclusion and Exclusion Criteria

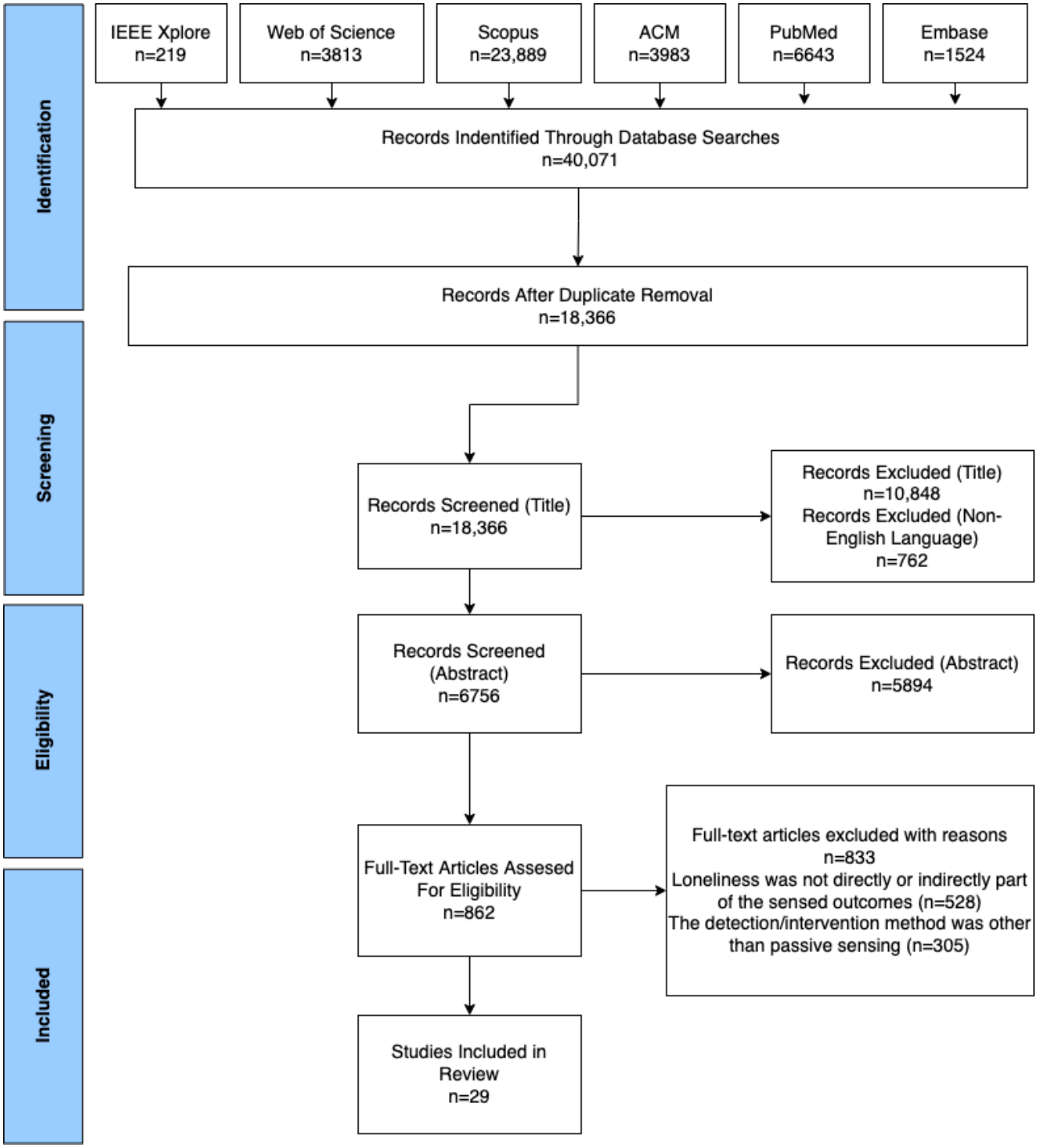
Studies were included if they were published in the English language and presented results on identifying loneliness and social isolation using passive sensing technology, such as smartphone apps, fitness trackers, and home sensors. Studies were excluded if they were written in a non-English language, if loneliness or social isolation was not one of the assessed outcomes, or if the detection method was other than passive sensing.

Data Extraction

The selection process is illustrated in Figure 1. A total of 40,071 studies were identified using the search terms. After title and abstract screening, 862 full texts were reviewed, of which 29 (3.4%) studies were selected for inclusion in the review, as follows:

- General description of the study: authors, year, and country
- Study design: population type and age, participant sample selection process, duration of the study, ground truth data collection methods, and privacy handling
- Study technology insights: technology used for sensing, data collection streams, and algorithm.
- Study outcome characteristics: indicators or identification markers and sensed outcomes.

Figure 1. Flow diagram of the literature search and selection process (adapted from PRISMA [Preferred Reporting Items for Systematic Reviews and Meta-Analyses]).



Results

Multimedia Appendix 1 [58-85] summarizes the main information from each study to address the questions under investigation.

Population

Most studies targeted young adults (10/29, 34%), most frequently college students [58-67], or older adults (12/29, 41%) [68-78,86], whereas 17% (5/29) of the studies included a mixed age group [79-84]. The studies sampled between 5 and 364 participants and lasted anywhere from 1 week to 5 years. In all, 3% (1/29) of studies was not longitudinal [61], and 7% (2/29) of other studies did not report the duration [75,81].

In all, 37% (11/29) of studies explicitly stated the participants' ages [61,62,64,67,68,70,75,76,78,83,84], whereas the other studies (18/29, 62%) mentioned only the age group without providing their age: younger adults [58-60,63,65,66], older adults [69,71-74,77,86], or mixed age groups [79-82]. Of the 11 studies, a total of 4 (36%) of studies focusing on a younger population reported an age range of 18 to 28 years [60-62,64], whereas 6 (55%) studies with a target sample of older adults reported an age range of 53 to 91 years [68,70,75,76,78,83]. Participants in 9% (1/11) of studies ranged in age from 18 to

78 years, representing a mixed group of younger and older adults [84].

With respect to the participants' gender, a total of 3 studies focusing on a younger population included both male and female participants [60,63,83]. Female participants dominated 3 studies [59,61,67], male participants dominated 2 others [64,66], and 1 study had exclusively female participants [62]. Female dominance has been clearly observed in older population studies. A total of 7 studies concerning older adults reported their gender, and all (7/29, 24%) these studies showed a female predominance, except one. The female ratios in these studies were 60% [70], 81% [72], 74% [68], 54% [74], 89% [75], 44% [78], and 88% [76].

A total of 11 studies mentioned participants' educational level [59,63-68,72,76,78,83]. The younger population mostly comprised college students [59,63-67]. In all, 2 studies recruited first-year college students [59,63], whereas 4 others targeting a younger population recruited a mix of university students at the graduate and undergraduate levels [61,65-67]. Only 4 studies reported the education levels of the older adults [68,72,76,78], of which 2 (50%) studies included participants who had completed their graduate studies [72,78], and the remaining 2 (50%) studies included participants with varying levels of education, with 46% (12/26) completing a college education in a study [76], 39% (19/50) completing primary level education, and 39% (19/50) having had no formal education in another study [68].

Participants' ethnic status varied. A total of 8 studies clearly reported the origin of their participants [59,60,63,65,66,68,75,78], 5 (63%) of which targeted a younger population from diverse ethnic groups [59,60,63,65,66], and 3 (37%) of which targeted older adults [68,75,78]. In the younger population, most participants had an Asian cultural background, whereas 2 studies were from the United States with participants from different backgrounds [59,66], 1 study was from China [65], and 1 was from Singapore [63]. The remainder were White, Hispanic, African American, or Caucasian [60,65]. A study from Singapore examining an older population had 87% Chinese participants [68], whereas another US-based study included 75% White participants [75].

Sensing Streams

Given that smartphones and wearable fitness trackers are equipped with a variety of physical sensors, they are highly effective instruments for monitoring users' movements and passively collecting data on their daily life patterns. Of the 29 studies, 18 (62%) used smartphones or wearables as a sensing modality to passively capture data [58-67,71,74,79,80,82-84,87].

Among the various physical sensors used in smartphones or wearable devices, the accelerometer and GPS are the most frequently used sensors for passive data collection. Indeed, 72% (13/18) of the studies that used a smartphone or smart watch as a sensing instrument made use of accelerometer and GPS sensors to gather data on a user's physical activities [59-63,65-67,74,80-83]. The accelerometer is used because of its low power use and low privacy concerns, whereas GPS is used to monitor the users' regularly visited locations and the time spent at those locations, which aids in determining their social habits. Another explanation for using GPS in several studies is that it is typically available with all smartphones.

Following this, Bluetooth and microphones are the second most widely used sensors in the identified studies, accounting for 39% (7/18) of studies that used passive sensing via smartphones or wearable devices to acquire user data [59,62,65-67,80]. Bluetooth is typically used as a form of proximity sensor to gather information about a user's sociability and physical encounters with other users who also have Bluetooth enabled in their devices, whereas a microphone is used to infer sleep and social experiences. In addition, 7% (2/29) of studies used Wi-Fi MAC addresses to collect information about users' sociability. Moreover, a study collected data from a smartphone light sensor to detect the screen lock or unlock status [82].

Apart from the sensors in smartphones, the studies included data on phone activity, which could also be used as an indicator of loneliness or social isolation. SMS text messages and call logs are the most frequently used; 56% (10/18) of the studies have used these logs to derive user communication activities [59,60,62-65,71,74,76,82]. In addition, 2 experiments integrated the number of times that each smartphone was locked and unlocked [59,82], and 2 studies gathered data on the types of apps that were used more often to ascertain users' mood and sociability levels [65,82]. To maximize learning opportunities regarding user habits, a study collected contact information, browsing history, and email-related data [64].

Of the 29 studies included here, 9 (31%) solely used ambient and other physical sensors installed at different locations in the home to learn about participants' in-home mobility patterns and behavior [68-70,72,73,75,77,81,86]. Different types of sensors were used, including passive infrared motion sensors installed on walls to collect motion data, pressure sensors to detect the presence of the participant on a bed or chair, sound sensors to detect social interactions and activities in the household, and door sensors to detect walking patterns and people movements.

Loneliness and Social Isolation Assessment

Different assessment methods have been used in experiments to validate the passively collected data. Most studies (21/29, 72%) relied on self-reports and questionnaires, which can be administered directly in a clinical setting or remotely using a mobile app. Self-reports or questionnaires were used to gather data on participants' physical and mental health in some of the included studies (8/29, 27%) that assessed social interactions and well-being [60,67,68,71,80,82-84], depression or anxiety [60,62,65,66,70,77,78], and daily activities [63,67,68,79,80,82]. Numerous scales have been used specifically for the detection of loneliness, including the UCLA loneliness scale [88], the De Jong Gierveld Loneliness Scale [41], and the ESTE-R Loneliness Scale [74]. The Social Interaction Anxiety Scale [89], Depression Anxiety Stress Scale [90], and Positive Affect Negative Affect Schedule [91] have been used to measure depression and anxiety symptoms in individuals. Some studies used self-developed questionnaires, such as the ecological momentary assessment, which were specifically related to the application domain and target populations.

Efficacy and Reliability

Detection Methods

Numerous markers were used in the included studies to identify changes in behavior, which may infer loneliness or social isolation. Of the various markers, the number of incoming and outgoing calls and SMS text messages is the most widely studied, with 24% (7/29) of studies () using this metric for detecting loneliness [60,63,64,72,74,76,82]. Similarly, time spent outside the home is also the most explored marker; 24% (7/29) of studies presented the relationship between time spent outside the home and its effect on loneliness levels in individuals [68,72-75,83,86]. Additional markers include sleep cycles that have been examined in 4 studies [59,66,68,70], voice activity and verbal communication explored by 4 studies [62,63,66,83], activity, mobility, and walking speed investigated in 6 studies [59,61,62,70,72,80], the average time spent in different areas of the household [68,71], and app types that have been studied by 2 studies [65,82]. Interaction with others detected through Bluetooth proximity sensing was examined in 5 studies [59,62,63,65,67], and 3 studies investigated the effect of speech activity on loneliness [62,63,83].

Phone Calls or SMS Text Messages

A total of 8 studies examined the relationship between incoming and outgoing phone calls and loneliness [64,65,71,72,74,76,82,84]. According to a study, individuals who received fewer incoming calls rated themselves as being more lonely on loneliness measurement scales compared with those who received more calls [82]. Another study found that loneliness was associated with fewer incoming calls but not with increased outgoing calls. The most lonely individual received only 40% of the number of calls received by the least lonely individual [76]. Another study explored the relationship between loneliness and calls from family members and friends. Outgoing and incoming family calls are significant characteristics of family and spousal loneliness, which is about feeling lonely within intimate relations. Similarly, for social loneliness, calls from friends were not considered a relevant attribute; rather, calls from acquaintances were considered a relevant attribute while developing predictive models for loneliness detection [74]. A total of 8 studies investigated the relationships between call frequency and call duration with loneliness. Individuals with greater loneliness received fewer calls on average according to a study [76]. There is a weak negative correlation between loneliness and call length, indicating that people who make longer calls have lower feelings of loneliness. In comparison, loneliness showed no significant relationship with the frequency of phone calls [65]. Another study found a similar negative correlation between phone conversation duration and social well-being [84]. In addition, 2 other studies found a similar negative correlation between the length of calls and loneliness on the UCLA loneliness scale [64,71]. Similarly, the frequency of phone calls was significantly linked to the length of time spent on the phone [72]. Another study found that students who received more phone calls, particularly on weekdays, reported reduced feelings of loneliness, better transition to college life, and a stronger sense of class community [63]. Regarding SMS text messages, individuals with a high score for social anxiety or loneliness rarely receive messages [82]. Another study discovered a weak negative correlation between loneliness and SMS text messages, particularly at night [65]. Although another study found no significant relationship between instant messenger and SMS text message frequency and loneliness or social well-being, phone call use was shown to be favorably associated with social well-being [84]. Another study discovered a significant correlation between the quantity of messages and UCLA loneliness scale [64].

Web Social Activity and Communication Apps

A total of 3 studies examined the relationship between web social engagement and loneliness [65,82,84]. According to a study conducted with a younger population, loneliness is positively linked to app use frequency, which implies that younger individuals who spend more time on social apps experience more loneliness [65]. However, a study targeting older adults found the opposite: older adults who spent more time on social media had substantially reduced feelings of loneliness [84]. The authors found that older adults' social media networks led to reduced feelings of loneliness and increased levels of general well-being compared with younger adults' social media networks [84]. Another study found that smartphone behaviors indicative of web social activity (eg, the frequency of using social or communication apps) were unrelated to social anxiety or loneliness [82].

Number of Computer Sessions or Hours on the Computer

In addition, 2 studies evaluated computer use in terms of the number of sessions and overall time spent on the computer daily [72,78]. They discovered that an increasing number of computer sessions was associated with higher levels of loneliness. Previous research on the relationship between computer use and loneliness has produced contradictory results. Although some argue that computer use helps combat loneliness [92], others argue that increasing computer use (particularly among young people) is associated with increased loneliness [93-96]. According to a study, telephone use and computer use were not shown to be substantially related to loneliness [72]. The authors stated that the lack of significance for these social factors may be explained in part by the strong correlation between them. That is, the number of phone calls was significantly associated with the amount of time spent on the phone, whereas the number of computer sessions was significantly associated with the amount of time spent on the computer [72].

Bluetooth Proximity Sensing

A total of 4 studies examined the association between Bluetooth proximity-sensing social interactions and feelings of loneliness [58,59,65,67]. According to a study, when a number of different Bluetooth devices are identified in the vicinity, people tend to report feelings of loneliness. This tendency may represent situations in which a person feels lonely in public places with a high concentration of Bluetooth devices [67]. Another study discovered a significant association between Bluetooth device encounters and the Patient Health Questionnaire-9 depression scale [58]. However, a study found no relationship between

Bluetooth proximity and loneliness [65]. A study used Bluetooth together with other sensor data, such as GPS, Wi-Fi, SMS text messages, and call logs, which showed better accuracy in predicting loneliness levels in individuals [59].

Daily Phone Use or Time on Mobile Phone

A total of 3 studies have explored the relationship between general smartphone use and loneliness [59,65,76]. According to a study, loneliness in older adults is associated with reduced daily phone use, to the point where the most lonely individual uses the phone almost two-thirds less than the least lonely one [76]. Another study found that decreased phone use during certain weekends and morning hours was associated with increased loneliness among the younger population [59]. Similarly, another study on students discovered that those at risk of loneliness spend more time on their smartphone each day and that loneliness was positively associated with app use, irrespective of the time of day [65].

Demographic Characteristics Affecting Loneliness

Of the 29 studies, 4 (14%) assessed gender differences in loneliness [76,82,84,87]. According to Petersen et al [76], gender had a substantial effect on the daily number of calls, with women making or receiving twice as many calls as men. Another study compared the gender of participants in high- and low-loneliness groups; the result of the chi-square test on the gender-based difference revealed no significance [82]. Another study found that women with a greater sense of well-being use reading apps and browsers more often than men do [87]. Similarly, 4 studies examined passive sensing and participants' behaviors in terms of age. A study assessed the ages of participants in high- and low-loneliness groups; the chi-square test indicated no statistical significance for the age difference [82]. According to another study, age was not a significant predictor of the daily number of calls [76]. In a study by Wetzel et al [84], the authors discovered a substantial relationship between participants' age and social media use time, suggesting that the association between social media use and perceived loneliness varies by age. This indicates that older adults who spend more time on social media experience substantially reduced feelings of loneliness. At younger ages, more time spent on social media was associated with increased levels of perceived loneliness. Another study found that the activity and frequency of messages and calls increased with the age of participants born before the year 2000 [79]. Cognitive ability was included as a demographic variable in 2 studies. According to a study, people with a chronic illness report significantly higher feelings of loneliness and poorer levels of social well-being than those without a chronic condition [84]. Another study discovered a relationship between better cognitive performance and higher daily call frequency [76]. In addition, this research examined whether an individual's pain level might serve as a predictor of phone calls but discovered no significant relationship between pain level and call frequency [76]. A total of 2 studies examined the relationship between an individual's personality traits and activity patterns inferred from passively collected smartphone data [64,79]. Moreover, a study discovered a significant correlation between emotional stability and extroverted personality characteristics and most smartphone-sensed features. In comparison, agreeableness, conscientiousness, and intellect personality traits are slightly associated with most of the features sensed by the smartphone, such as the number of messages, number of browser searches, number of calls, number of long incoming or outgoing calls, and number of contacts [64]. According to another study, gratitude is associated with participants' message and call patterns, with the most grateful participants communicating mostly through their smartphones [79]. Regarding occupational status, participants who are retired, are homemakers, or are unemployed are typically the least active and more prone to loneliness, whereas those enrolled in college or working full-time or self-employed are typically the most active with lower levels of loneliness. This analysis was performed using smartphone sensor data collected from an accelerometer, GPS, microphone, SMS text messages, and call logs [79]. Another study that used smartphone app communication data as an indicator of loneliness across the adult life span during the COVID-19 pandemic found that individuals who did not have a partner reported greater feelings of loneliness and poorer social well-being than those who had a partner [84]. The reason for this might be that during the COVID-19 pandemic, everyone was required to stay at home, and those without a partner were more prone to experiencing loneliness than those in a relationship.

Time Outside of Home and Time Spent at Home

A total of 7 studies examined the effect of time spent outdoors or at home on an individual's loneliness levels [68,72,73,75,77,83,86]. A study with older adults found that spending more time away from home was associated with decreased levels of loneliness [75]. Similarly, other research with both young and older participants discovered that people who spent more time outside the home reported fewer feelings of loneliness [83]. Other research has analyzed older adults' outdoor trips using two measures: average daily outdoor time and number of outdoor visits. They discovered that spending more time outdoors resulted in decreased levels of loneliness and higher social networking scores. In addition, 2 other studies discovered a strong association between the daily time spent outside the house and loneliness [72,77]. Similarly, another study discovered that individuals who spend more time at home are lonely. The physical activity score is positively associated with the average amount of time spent outside the house, suggesting that outings are also an important indicator of physical activity [86]. A preliminary analysis found no relationship between the time spent outside the home and loneliness [73]. They stated that their participants were more technology literate and may have engaged in social activities such as computer or mobile use that do not require leaving the home.

Daily Activity and Movement

A total of 5 studies investigated the effects of daily activities and movements and their duration on loneliness levels [59,65,66,68,83]. A study discovered that individuals performing more physical activities experience less loneliness [66]. Similarly, greater dispositional loneliness (UCLA loneliness scale) was associated with a substantially shorter mean movement duration in another study [83]; dispositional

loneliness is defined by a sense of disconnection from others and distressing emotions of isolation [97]. In addition, loneliness increased with the number of significant places visited [83]. Loneliness is also negatively correlated with movement duration [83]. Other research conducted with students discovered a strong negative correlation between loneliness and indoor mobility; for example, the amount of time a student moves indoors throughout the day. Research indicates that inactive students are more likely to experience loneliness. Another study found that people who spend more time engaging in activities are less lonely regardless of the time of day or night [65]. Another study found a similar pattern: a higher total amount of movement and steps throughout the day and night time hours results in less feelings of loneliness [59]. Deviation from one's regular geospatial activity was associated with a substantial reduction in daily stress and loneliness [68].

Conversation Activity

A total of 3 studies investigated the effect of conversation and its duration on loneliness levels [58,66,83]. In addition, 2 studies found no significant correlation between speech duration and loneliness [66,83]. However, another study found that students who engage in less conversational contacts are more likely to experience depression [58].

Sleep

A total of 3 studies investigated the relationship between sleep duration and feelings of loneliness and stress [58,66,68]. According to a study, sleep duration was not correlated with loneliness but was negatively related to daily stress, meaning that people who get enough sleep experience less stress [66]. Another study with students discovered that those who sleep fewer hours are more prone to depression [58]. A study focusing on older adults found that those who perceived themselves as socially isolated had more midday naps [68].

Privacy Issues and Ethical Concerns

Overall, 69% (20/29) of the studies did not address privacy or ethical issues and did not explicitly mention privacy concerns [59,64-69,71-83]. A total of 5 studies provided details of ethical approval from the relevant committees and participant consent [60-62,70,84]. In addition, 2 studies stated that they maintained user data on a secure data server and anonymized user identifiers to protect users' privacy [58,85]. Moreover, a study collected statistical data on smartphone use only to preserve participants' privacy [63]. Another study, which uses a video camera placed in a smart home to monitor users' activities, reported that they captured footage for only 5 seconds whenever a motion was detected at the door [86].

Discussion

Comparison With Previous Work

Numerous review studies have examined the use of passive sensing to monitor a variety of mental and physical health and well-being outcomes. Many aspects of passive sensing have been covered in previous review studies, providing researchers with current challenges and suggestions for potential future research. These reviews have focused on different mental and physical health conditions, such as stress [98], mood disorders [99], sleep problems [100], cardiac issues [101], chronic health conditions in older adults [102] and schizophrenia [103]. Compared with previous reviews, this review is the first to focus on the use of passive sensing for loneliness and social isolation detection and to explore the limitations of the passive sensing systems used.

Findings From Reviewed Studies

The targeted population is one of the key dimensions to be considered when designing and developing loneliness detection systems. Most studies (17/29, 58%) included in this scoping review targeted younger populations. It may be relatively easier to generate a sample group of younger people from colleges or universities, as opposed to older adults, who are dispersed across community settings and may not be readily accessible through services, education, or workspaces. Evidence in the literature suggests that isolation is typically higher in late adolescence and after retirement than in the middle age range [14]. Moreover, unemployed individuals in the middle age group are more susceptible to loneliness [104]. An age-related association was observed in the detection methods; loneliness in younger adults was assessed mostly through smartphones [58-60,62,63,65-67], whereas ambient sensors were used more frequently by older adults [68-78,86]. Ambient sensing may be used by older people because they are less likely to use wearables and have higher privacy concerns [105]. Some older adults may not be able to use smartphones because of lack of technological literacy and complex user interfaces [106]. In addition, people with dementia can forget to carry or wear such devices or charge their smartphones in a timely manner, resulting in intermittent or no data collection. Furthermore, there is a strong possibility that some wearable sensors create an unpleasant feeling during prolonged skin connection (eg, electrodes on the skin). As a result, older adults may reject such wearables, particularly at home [107].

Before they are widely available to the public, these loneliness detection systems must undergo extensive validation to ensure that users feel comfortable using and trusting them. The validation of such detection systems is important because it provides relevant input and knowledge about the performance and reliability of the systems to researchers and developers. Two facets of validation of these loneliness detection systems should be discussed: participant selection and technology's efficacy in assessing social outcomes. With regard to population selection, most of the included studies (17/29, 58%) recruited younger adults and college students via advertisements on student mailing lists or Facebook groups. Moreover, most of the studies recruited first-year college students or undergraduate students from the same college or university, with a random ratio of male and female participants. A study recruited students from a specific class only [58], resulting in a population with very similar sociodemographic backgrounds, daily routines, and study challenges. There is a lack of detail on randomization in student population selection, either in terms of sample collection from more than one college or from younger populations other than college students. A study recruited international students because they were considered at a higher risk of loneliness [64]. In addition, many studies (9/29, 31%)

did not include sufficient information on the participants' socioeconomic status. In the case of older adults, studies recruited participants who resided in sensor-enabled homes created by specific senior citizen welfare organizations. Some studies (6/29, 20%) omitted older adults with physical, sensory, or cognitive impairment, which may have resulted in demographic selection bias because people with disabilities may experience greater loneliness [108,109]. Most studies (9/12, 75%) have focused on older adults who were already living alone to assess their loneliness, which may result in biased findings. Few studies (3/29, 10%) used the Mini-Mental State Examination [110] to screen and exclude older adults with cognitive impairment. In addition, some studies (7/29, 24%) have not mentioned their sample selection or recruitment process for both the younger and older populations; instead, they simply reported the number of individuals from a specific age group.

Most studies (24/29, 82%) have used supervised machine learning techniques to infer loneliness levels from passively acquired data. In supervised machine learning, labeled data are required to train a model, where the gathered ground truth data are used to label the raw sensor data. Ground truth data have been used to compare and validate passively acquired data through smartphones and other passive devices. Inaccuracies in the ground truth evidence result in incorrect model training and an inefficient detection system. Thus, ground truth data were critical in these studies. Ground truth data from assessments, such as loneliness assessment questionnaires, are used to label the data collected through passive sensing streams with the corresponding ground truth state. Most studies (21/29, 72%) have used self-reports and questionnaires, which can then be used for validation purposes with passively collected data. This method has some drawbacks, such as users may not always respond to such questionnaires, or there could be accuracy issues depending on the participants' current mental and physical state.

Participants' motives, preferences, and expectations regarding surveillance systems can influence their willingness to use available solutions. Some of those interests could be physiological, such as improving behaviors in daily life, and some could be more technical, such as system scalability and acceptability. Most studies (18/29, 62%) that use smartphone apps to passively acquire user data have used Android-based mobile apps because of their lower privacy restrictions compared with the iPhone iOS operating system when collecting data through various sensors available on smartphones. Another reason might be cost, as Android devices are more affordable and are the most widely used operating system, encouraging more users to use the proposed systems.

Another technological consideration of smartphone-based detection systems is their power use. Owing to continuous data collection and constant use of energy by various sensors, the battery life of smartphones degrades more quickly, which is a major concern for users. Only a few studies (6/29, 20%) have addressed this topic or plan to do so in the future. The approach used to reduce power use is to reduce the sampling rate of sensors when the battery level falls below a certain threshold or to use less beneficial sensors only once a day [48].

Apart from energy use, the participants' primary concern is privacy [48]. Although most of the reports did not mention privacy concerns, few studies (8/29, 28%) indicated that this is one of the key concerns of participants. These studies indicated that they obtained approval from participants for passive data collection or that they obtained privacy and ethical clearance from an organization. Several (6/8, 75%) of these studies used anonymous participant identifiers to store data and transmitted them securely to the server. However, a sizable portion of the studies did not address privacy concerns, which raises questions about their acceptability to participants and the wider public and about the potential harms for this detection method in the case of data and privacy breaches.

Implications for Future Research

This review demonstrates that previous research has not drawn strong distinctions between social isolation and loneliness, with many researchers using both terms interchangeably. According to previous research, loneliness and isolation overlap relatively little, with most lonely times occurring when not alone, for example, when an individual is in a crowd of strangers [67]. Across all self-reports, loneliness seems to be greatest when participants also report being alone at the time and lowest when participants additionally report being with a significant other (partner or friend) [67]. In addition, some experiments have intertwined loneliness with other related concepts. For instance, in a study [111], the authors recorded the number of visits to older adults living alone and did not attempt to explicitly detect loneliness. Thus, it is necessary to distinguish loneliness from other similar terms, most notably, social isolation.

Another significant gap pertains to privacy concerns. Most research has not focused on privacy issues. A few studies (9/29, 31%) discussed their strategies for maintaining device and data privacy, which included anonymization and safe data transmission to data collection servers; however, most of the studies (18/29, 62%) plainly stated that they gathered data with the consent of the users but did not discuss data privacy and security. Most studies (12/18, 66%) collecting participant data passively through smartphones and wearables were based in the United States, China, and Asia, with very few studies conducted in European countries. This may be because the European Union's General Data Protection Regulations place stringent restrictions on data collection and transfer. It is important to engage potential study participants in the design, development, and validation phases of such systems to ensure that the system satisfies their expectations [112]. This is crucial if the system is to be adopted extensively and sustainably by the target population.

Many smartphone or wearable users prefer tracking systems to provide them with valuable knowledge and feedback about their activities and well-being status [113]. When people receive well-being-based feedback, they are more likely to make positive changes in their lifestyle and behaviors related to physical activity, including well-being, sociability, and mental health [114,115]. However, little is known about the impact of user feedback on behavioral changes related to loneliness. As demonstrated in this scoping review, existing systems are deficient in delivering real-time feedback to participants, which

can help participants develop interest and trust in such tracking systems and help them to make attempts to change their lifestyle in response to system recommendations. Experts are examining ways to develop these tracking systems to provide valuable feedback, alerts, and advice to users to improve their current mental health [116].

Most of the studies (25/29, 86%) in this scoping review were pilot or feasibility studies with very small or limited sample sizes. The validation of these detection systems is a crucial component, and the systems should be validated with a subset of the target population that is highly representative over an adequate period to obtain appropriate data to deliver as reliable results as possible. Many studies are of short duration, which may provide an insufficient window for detection and build participants' trust and acceptability in detection systems. In addition, some of the proposed systems selected population samples from a very specific population type or at a specific period, such as a lockdown during a pandemic crisis [84], which may result in misleading outcomes when applied to other populations or at different periods. Similarly, the population selection methods used were not rigorous, resulting in potential self-selection bias and limited generalizability, particularly with low representations of at-risk groups, such as unemployed people or older adults not in supported housing.

Limitations

Although we conducted a thorough search of computer science, health, and social science databases using multiple search terms related to loneliness and social isolation detection, some articles could have been overlooked. For example, studies that focused on general mental health or emotion recognition, with loneliness or social isolation serving as a subset of larger research, were not included. Furthermore, because we included only studies conducted in English, there might be several studies on loneliness detection published in other languages that were not included in this scoping analysis.

Conclusions

It is evident that the use of smartphones, wearable smart devices, and ambient sensors to detect loneliness and social isolation in different age groups has increased in the last few years. Compared with more conventional tracking systems, smartphones are simple to use, unobtrusive, familiar, and inexpensive. They also have a variety of sensors that allow the collection of users' data in real time, without interfering with users' daily activities. This comprehensive scoping review reveals that smartphones and mobile and ambient sensing systems have the potential to monitor users' behaviors and daily activities to infer loneliness and social isolation, and it is likely that research interest in this field will grow in the future. However, most existing methods have shortcomings, particularly in privacy preservation and validation across diverse populations, which need to be rigorously addressed in future research. Finally, it is worth noting that researchers need to investigate what motivates people to use such tracking mechanisms and what inspires their trust and long-term adherence if they are to be adopted and implemented within wider populations.

Acknowledgments

This work has been funded by SFI Centre for Research Training in Advanced Networks for Sustainable Societies (ADVANCE CRT), which is a part of Science Foundation Ireland.

Authors' Contributions

MMQ worked on the study design, database searching, filtering, data analysis, and drafting. EZ worked on the study design, database searching, filtering, data analysis, and critical revision of the article. DP worked on the conceptualization, study design, analysis, and critical revision of the article. EBW worked on the conceptualization, study design, analysis, and critical revision of the article.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The characteristics of the included studies.
[DOCX File , 35 KB - mhealth_v10i4e34638_app1.docx]

References

1. Baumeister RF, Leary MR. The need to belong: desire for interpersonal attachments as a fundamental human motivation. Psychol Bulletin 1995;117(3):497-529. [doi: 10.1037/0033-2909.117.3.497]
2. Snijders T, Bosker R. The worship and principles of the Church of England: a sermon, preached at the opening of Christ's Church, in Great-Barrington, on Christ-Mass Day, M.DCC.LXIV. / By Thomas Davies, A.M. and missionary from the society. -: William Goddard; 1999.
3. Campbell M. Loneliness, social anxiety and bullying victimization in young people: a literature review. Psychol Educ 2013;50:567 [FREE Full text]

4. Martín-María N, Caballero FF, Miret M, Tyrovolas S, Haro JM, Ayuso-Mateos JL, et al. Differential impact of transient and chronic loneliness on health status. A longitudinal study. Psychol Health 2020 Feb 25;35(2):177-195. [doi: 10.1080/08870446.2019.1632312] [Medline: 31237442]
5. Measuring national well-being: our relationships. Office for National Statistics. 2015. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/articles/measuringnationalwellbeing/2015-02-12 [accessed 2022-04-05]
6. Social isolation: a conceptual and measurement proposal. University of Oxford. 2014. URL: https://opendocs.ids.ac.uk/opendocs/handle/20.500.12413/11811 [accessed 2022-04-05]
7. Grenade L, Boldy D. Social isolation and loneliness among older people: issues and future challenges in community and residential settings. Aust Health Rev 2008 Aug;32(3):468-478. [doi: 10.1071/ah080468] [Medline: 18666874]
8. de Jong Gierveld J, Havens B. Cross-national comparisons of social isolation and loneliness: introduction and overview. Can J Aging 2004 Mar 31;23(2):109-113. [doi: 10.1353/cja.2004.0021] [Medline: 15334811]
9. Cacioppo JT, Hawkley LC, Berntson GG. The anatomy of loneliness. Curr Dir Psychol Sci 2016 Jun 23;12(3):71-74. [doi: 10.1111/1467-8721.01232]
10. Sadler WA, Weiss RS. Loneliness: the experience of emotional and social isolation. Contemporary Sociol 1975 Mar;4(2):171. [doi: 10.2307/2062224]
11. Heinrich LM, Gullone E. The clinical significance of loneliness: a literature review. Clin Psychol Rev 2006 Oct;26(6):695-718. [doi: 10.1016/j.cpr.2006.04.002] [Medline: 16952717]
12. Routasalo P, Pitkala KH. Loneliness among older people. Rev Clin Gerontol 2004 Aug 11;13(4):303-311. [doi: 10.1017/s095925980400111x]
13. Cacioppo JT, Hawkley LC, Ernst JM, Burleson M, Berntson GG, Nouriani B, et al. Loneliness within a nomological net: an evolutionary perspective. J Res Personality 2006 Dec;40(6):1054-1085. [doi: 10.1016/j.jrp.2005.11.007]
14. Luhmann M, Hawkley LC. Age differences in loneliness from late adolescence to oldest old age. Dev Psychol 2016 Jun;52(6):943-959 [FREE Full text] [doi: 10.1037/dev0000117] [Medline: 27148782]
15. Pinquart M, Sorensen S. Influences on loneliness in older adults: a meta-analysis. Basic Appl Soc Psychol 2010 Jun 07;23(4):245-266. [doi: 10.1207/s15324834basp2304_2]
16. Park NS, Jang Y, Lee BS, Chiriboga DA. The relation between living alone and depressive symptoms in older Korean Americans: do feelings of loneliness mediate? Aging Ment Health 2017 Mar 14;21(3):304-312. [doi: 10.1080/13607863.2015.1099035] [Medline: 26465083]
17. Hawkley LC, Thisted RA, Masi CM, Cacioppo JT. Loneliness predicts increased blood pressure: 5-year cross-lagged analyses in middle-aged and older adults. Psychol Aging 2010 Mar;25(1):132-141 [FREE Full text] [doi: 10.1037/a0017805] [Medline: 20230134]
18. Henriksen J, Larsen ER, Mattisson C, Andersson NW. Loneliness, health and mortality. Epidemiol Psychiatr Sci 2017 Oct 30;28(2):234-239. [doi: 10.1017/s2045796017000580]
19. Valtorta NK, Kanaan M, Gilbody S, Ronzi S, Hanratty B. Loneliness and social isolation as risk factors for coronary heart disease and stroke: systematic review and meta-analysis of longitudinal observational studies. Heart 2016 Jul 01;102(13):1009-1016 [FREE Full text] [doi: 10.1136/heartjnl-2015-308790] [Medline: 27091846]
20. Uchino BN, Uno D, Holt-Lunstad J. Social support, physiological processes, and health. Curr Dir Psychol Sci 2016 Jun 23;8(5):145-148. [doi: 10.1111/1467-8721.00034]
21. Cacioppo JT, Ernst JM, Burleson MH, McClintock MK, Malarkey WB, Hawkley LC, et al. Lonely traits and concomitant physiological processes: the MacArthur social neuroscience studies. Int J Psychophysiol 2000 Mar;35(2-3):143-154. [doi: 10.1016/s0167-8760(99)00049-5]
22. Victor CR, Yang K. The prevalence of loneliness among adults: a case study of the United Kingdom. J Psychol 2012;146(1-2):85-104. [doi: 10.1080/00223980.2011.613875] [Medline: 22303614]
23. Häfner S, Baumert J, Emeny R, Lacruz M, Bidlingmaier M, Reincke M, MONICA/KORA Study Investigators. To live alone and to be depressed, an alarming combination for the renin-angiotensin-aldosterone-system (RAAS). Psychoneuroendocrinology 2012 Feb;37(2):230-237. [doi: 10.1016/j.psyneuen.2011.06.007] [Medline: 21742440]
24. Mehdi M, Stach M, Riha C, Neff P, Dode A, Pryss R, et al. Smartphone and mobile health apps for tinnitus: systematic identification, analysis, and assessment. JMIR Mhealth Uhealth 2020 Aug 18;8(8):e21767 [FREE Full text] [doi: 10.2196/21767] [Medline: 32808939]
25. Enck W, Gilbert P, Han S, Tendulkar V, Chun B, Cox LP, et al. TaintDroid: an information-flow tracking system for realtime privacy monitoring on smartphones. ACM Transact Comput Syst 2014 Jun;32(2):1-29. [doi: 10.1145/2619091]
26. Cheffena M. Fall detection using smartphone audio features. IEEE J Biomed Health Inform 2016 Jul;20(4):1073-1080. [doi: 10.1109/JBHI.2015.2425932] [Medline: 25915965]
27. Abowd GD, Mynatt ED. Charting past, present, and future research in ubiquitous computing. ACM Transact Comput Human Interact 2000 Mar;7(1):29-58. [doi: 10.1145/344949.344988]
28. Cornwell EY, Waite LJ. Social disconnectedness, perceived isolation, and health among older adults. J Health Soc Behav 2009 Mar 01;50(1):31-48 [FREE Full text] [doi: 10.1177/002214650905000103] [Medline: 19413133]
29. Rashidi P, Mihailidis A. A survey on ambient-assisted living tools for older adults. IEEE J Biomed Health Inform 2013 May;17(3):579-590. [doi: 10.1109/jbhi.2012.2234129] [Medline: 24592460]

30. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Oct 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
31. Mays N, Pope C, Popay J. Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field. J Health Serv Res Policy 2005 Jul;10 Suppl 1:6-20. [doi: 10.1258/1355819054308576] [Medline: 16053580]
32. O'Brien K, Wilkins A, Zack E, Solomon P. Scoping the field: identifying key research priorities in HIV and rehabilitation. AIDS Behav 2010 May 10;14(2):448-458. [doi: 10.1007/s10461-009-9528-z] [Medline: 19277858]
33. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
34. Russell DW. UCLA Loneliness Scale (Version 3): reliability, validity, and factor structure. J Pers Assess 1996 Mar;66(1):20-40. [doi: 10.1207/s15327752jpa6601_2] [Medline: 8576833]
35. VanderWeele TJ, Hawkley LC, Thisted RA, Cacioppo JT. A marginal structural model analysis for loneliness: implications for intervention trials and clinical practice. J Consult Clin Psychol 2011 May;79(2):225-235 [FREE Full text] [doi: 10.1037/a0022610] [Medline: 21443322]
36. Townley G, Kloos B. Mind over matter? The role of individual perceptions in understanding the social ecology of housing environments for individuals with psychiatric disabilities. Am J Community Psychol 2014 Dec;54(3-4):205-218. [doi: 10.1007/s10464-014-9664-0] [Medline: 24917220]
37. Hays R, DiMatteo MR. A short-form measure of loneliness. J Pers Assess 1987 Mar 1;51(1):69-81. [doi: 10.1207/s15327752jpa5101_6] [Medline: 3572711]
38. de Jong-Gierveld J, Kamphuls F. The development of a rasch-type loneliness scale. Applied Psychol Measure 2016 Jul 27;9(3):289-299. [doi: 10.1177/014662168500900307]
39. Manual of the Loneliness Scale 1999. Amsterdam, Netherlands: Department of Social Research Methodology, Vrije Universiteit; 1999.
40. Gierveld JD, Tilburg TV. A 6-item scale for overall, emotional, and social loneliness. Res Aging 2016 Aug 18;28(5):582-598. [doi: 10.1177/0164027506289723]
41. De Jong Gierveld J, Van Tilburg T. The De Jong Gierveld short scales for emotional and social loneliness: tested on data from 7 countries in the UN generations and gender surveys. Eur J Ageing 2010 Jul;7(2):121-130 [FREE Full text] [doi: 10.1007/s10433-010-0144-6] [Medline: 20730083]
42. Berkman L, Syme S. Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda county residents. Am J Epidemiol 1979 Mar;109(2):186-204. [doi: 10.1093/oxfordjournals.aje.a112674] [Medline: 425958]
43. Koenig HG, Westlund RE, George LK, Hughes DC, Blazer DG, Hybels C. Abbreviating the duke social support index for use in chronically ill elderly individuals. Psychosomatics 1993 Jan;34(1):61-69. [doi: 10.1016/s0033-3182(93)71928-3]
44. Pachana NA, Smith N, Watson M, McLaughlin D, Dobson A. Responsiveness of the Duke Social Support sub-scales in older women. Age Ageing 2008 Nov 01;37(6):666-672. [doi: 10.1093/ageing/afn205] [Medline: 18829690]
45. Lubben JE. Assessing social networks among elderly populations. Fam Community Health 1988;11(3):42-52. [doi: 10.1097/00003727-198811000-00008]
46. Steptoe A, Shankar A, Demakakos P, Wardle J. Social isolation, loneliness, and all-cause mortality in older men and women. Proc Natl Acad Sci U S A 2013 Apr 09;110(15):5797-5801 [FREE Full text] [doi: 10.1073/pnas.1219686110] [Medline: 23530191]
47. Russell D. The measurement of loneliness. In: Loneliness: A Sourcebook of Current Theory, Research and Therapy. United States: Wiley; 1982.
48. Trifan A, Oliveira M, Oliveira JL. Passive sensing of health outcomes through smartphones: systematic review of current solutions and possible limitations. JMIR Mhealth Uhealth 2019 Aug 23;7(8):e12649 [FREE Full text] [doi: 10.2196/12649] [Medline: 31444874]
49. Heath K, Guibas L. Multi-person tracking from sparse 3D trajectories in a camera sensor network. In: Proceedings of the 2008 Second ACM/IEEE International Conference on Distributed Smart Cameras. 2008 Presented at: 2008 Second ACM/IEEE International Conference on Distributed Smart Cameras; Sep 7-11, 2008; Palo Alto, CA, USA. [doi: 10.1109/icdsc.2008.4635679]
50. Wang F, Stone E, Skubic M, Keller JM, Abbott C, Rantz M. Toward a passive low-cost in-home gait assessment system for older adults. IEEE J Biomed Health Inform 2013 Mar;17(2):346-355. [doi: 10.1109/jbhi.2012.2233745]
51. Hayes T, Pavel M, Larimer N, Tsay I, Nutt J, Adami A. Distributed healthcare: simultaneous assessment of multiple individuals. IEEE Pervasive Comput 2007 Jan;6(1):36-43. [doi: 10.1109/mprv.2007.9]
52. Smith JR, Fishkin KP, Jiang B, Mamishev A, Philipose M, Rea AD, et al. RFID-based techniques for human-activity detection. Commun ACM 2005 Sep;48(9):39-44. [doi: 10.1145/1081992.1082018]
53. Madan A, Cebrian M, Lazer D, Pentland A. Social sensing for epidemiological behavior change. In: Proceedings of the 12th ACM international conference on Ubiquitous computing. 2010 Presented at: Ubicomp '10: The 2010 ACM Conference on Ubiquitous Computing; Sep 26 - 29, 2010; Copenhagen Denmark. [doi: 10.1145/1864349.1864394]
54. Golder SA, Macy MW. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. Science 2011 Sep 30;333(6051):1878-1881. [doi: 10.1126/science.1202775] [Medline: 21960633]

55. Farrahi K, Gatica-Perez D. What did you do today?: discovering daily routines from large-scale mobile data. In: Proceedings of the 16th ACM international conference on Multimedia. 2008 Presented at: MM08: ACM Multimedia Conference 2008; Oct 26 - 31, 2008; Vancouver British Columbia Canada. [doi: 10.1145/1459359.1459503]
56. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol 2009 Oct;62(10):e1-34 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.06.006] [Medline: 19631507]
57. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Sep 04;169(7):467. [doi: 10.7326/M18-0850]
58. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2014 Presented at: UbiComp '14: The 2014 ACM Conference on Ubiquitous Computing; Sep 13 - 17, 2014; Seattle Washington. [doi: 10.1145/2632048.2632054]
59. Doryab A, Villalba DK, Chikersal P, Dutcher JM, Tumminia M, Liu X, et al. Identifying behavioral phenotypes of loneliness and social isolation with passive sensing: statistical analysis, data mining and machine learning of smartphone and Fitbit data. JMIR Mhealth Uhealth 2019 Jul 24;7(7):e13209 [FREE Full text] [doi: 10.2196/13209] [Medline: 31342903]
60. Jacobson NC, Summers B, Wilhelm S. Digital biomarkers of social anxiety severity: digital phenotyping using passive smartphone sensors. J Med Internet Res 2020 May 29;22(5):e16875 [FREE Full text] [doi: 10.2196/16875] [Medline: 32348284]
61. Quiroz JC, Geangu E, Yong MH. Emotion recognition using smart watch sensor data: mixed-design study. JMIR Ment Health 2018 Aug 08;5(3):e10153 [FREE Full text] [doi: 10.2196/10153] [Medline: 30089610]
62. Poudyal A, van Heerden A, Hagaman A, Maharjan SM, Byanjankar P, Subba P, et al. Wearable digital sensors to identify risks of postpartum depression and personalize psychological treatment for adolescent mothers: protocol for a mixed methods exploratory study in rural Nepal. JMIR Res Protoc 2019 Sep 11;8(8):e14734 [FREE Full text] [doi: 10.2196/14734] [Medline: 31512581]
63. Yow W, Li X, Hung W, Goldring M, Cheng L, Gu Y. Predicting social networks and psychological outcomes through mobile phone sensing. In: Proceedings of the 2014 IEEE International Conference on Communications (ICC). 2014 Presented at: 2014 IEEE International Conference on Communications (ICC); Jun 10-14, 2014; Sydney, Australia. [doi: 10.1109/icc.2014.6883934]
64. Pulekar G, Agu E. Autonomously sensing loneliness and its interactions with personality traits using smartphones. In: Proceedings of the 2016 IEEE Healthcare Innovation Point-Of-Care Technologies Conference (HI-POCT). 2016 Presented at: 2016 IEEE Healthcare Innovation Point-Of-Care Technologies Conference (HI-POCT); Nov 9-11, 2016; Cancun, Mexico. [doi: 10.1109/hic.2016.7797715]
65. Li Z, Shi D, Wang F, Liu F. Loneliness Recognition Based on Mobile Phone Data. Paris: Atlantis Press; 2016.
66. Ben-Zeev D, Scherer EA, Wang R, Xie H, Campbell AT. Next-generation psychiatric assessment: using smartphone sensors to monitor behavior and mental health. Psychiatr Rehabil J 2015 Sep;38(3):218-226 [FREE Full text] [doi: 10.1037/prj0000130] [Medline: 25844912]
67. Wu C, Barczyk AN, Craddock RC, Harari GM, Thomaz E, Shumake JD, et al. Improving prediction of real-time loneliness and companionship type using geosocial features of personal smartphone data. Smart Health 2021 Apr;20:100180. [doi: 10.1016/j.smhl.2021.100180]
68. Sensor-driven detection of social isolation in community-dwelling elderly. In: Human Aspects of IT for the Aged Population. Cham: Springer; 2017.
69. Hu R, Pham H, Buluschek P, Gatica-Perez D. Elderly people living alone: detecting home visits with ambient and wearable sensing. In: Proceedings of the 2nd International Workshop on Multimedia for Personal Health and Health Care. 2017 Presented at: MM '17: ACM Multimedia Conference; Oct 23, 2017; Mountain View California. [doi: 10.1145/3132635.3132649]
70. Walsh L, Kealy A, Loane J, Doyle J, Bond R. Inferring health metrics from ambient smart home data. In: Proceedings of the 2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2014 Presented at: 2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); Nov 2-5, 20145; Houston, TX, USA. [doi: 10.1109/bibm.2014.6999237]
71. Martinez A, Ortiz V, Estrada H, Gonzalez M. A predictive model for automatic detection of social isolation in older adults. In: Proceedings of the 2017 International Conference on Intelligent Environments (IE). 2017 Presented at: 2017 International Conference on Intelligent Environments (IE); Aug 21-25, 2017; Seoul, Korea. [doi: 10.1109/ie.2017.35]
72. Austin J, Dodge HH, Riley T, Jacobs PG, Thielke S, Kaye J. A smart-home system to unobtrusively and continuously assess loneliness in older adults. IEEE J Transl Eng Health Med 2016;4:1-11. [doi: 10.1109/jtehm.2016.2579638]
73. Ambient sensor monitoring: detecting loneliness and depression in older adults via passive home sensing. University of South Florida. URL: https://www.academia.edu/31637787/Ambient_Sensor_Monitoring_Detecting_Loneliness_and_Depression_in_Older_Adults_via_Passive_Home_Sensing [accessed 2022-04-05]

74. Sanchez W, Martinez A, Campos W, Estrada H, Pelechano V. Inferring loneliness levels in older adults from smartphones. J Ambient Intell Smart Environ 2015;7(1):85-98. [doi: 10.3233/AIS-140297]
75. Lyons BE, Austin D, Seelye A, Petersen J, Yeargers J, Riley T, et al. Pervasive computing technologies to continuously assess Alzheimer's disease progression and intervention efficacy. Front Aging Neurosci 2015 Jun 10;7:102 [FREE Full text] [doi: 10.3389/fnagi.2015.00102] [Medline: 26113819]
76. Petersen J, Thielke S, Austin D, Kaye J. Phone behaviour and its relationship to loneliness in older adults. Aging Ment Health 2016 Oct;20(10):1084-1091 [FREE Full text] [doi: 10.1080/13607863.2015.1060947] [Medline: 26133745]
77. Huynh S, Tan H, Lee Y. Towards unobtrusive mental well-being monitoring for independent-living elderly. In: Proceedings of the 4th International on Workshop on Physical Analytics. 2017 Presented at: MobiSys'17: The 15th Annual International Conference on Mobile Systems, Applications, and Services; Jun 19, 2017; Niagara Falls New York USA. [doi: 10.1145/3092305.3092311]
78. Leese M, Bernstein J, Dorociak K, Mattek N, Wu C, Beattie Z, et al. Older adults' daily activity and mood changes detected during the COVID-19 pandemic using remote unobtrusive monitoring technologies. Innov Aging 2021;5(4):igab032-igab331 [FREE Full text] [doi: 10.1093/geroni/igab032] [Medline: 34671706]
79. Servia-Rodríguez S, Rachuri K, Mascolo C, Rentfrow P, Lathia N, Sandstrom G. Mobile sensing at the service of mental well-being: a large-scale longitudinal study. In: Proceedings of the 26th International Conference on World Wide Web. 2017 Presented at: WWW '17: 26th International World Wide Web Conference; April 3 7, 2017; Perth Australia. [doi: 10.1145/3038912.3052618]
80. Gu F, Niu J, He Z, Jin X, Rodrigues J. Smartbuddy: an integrated mobile sensing and detecting system for family activities. In: Proceedings of the GLOBECOM 2017 - 2017 IEEE Global Communications Conference. 2017 Presented at: GLOBECOM 2017 - 2017 IEEE Global Communications Conference; Dec 4-8, 2017; Singapore. [doi: 10.1109/glocom.2017.8254140]
81. Costa A, Rincon JA, Carrascosa C, Julian V, Novais P. Emotions detection on an ambient intelligent system using wearable devices. Future Generation Comput Syst 2019 Mar;92:479-489. [doi: 10.1016/j.future.2018.03.038]
82. Gao Y, Li A, Zhu T, Liu X, Liu X. How smartphone usage correlates with social anxiety and loneliness. PeerJ 2016;4:e2197-e2203 [FREE Full text] [doi: 10.7717/peerj.2197] [Medline: 27478700]
83. Fulford D, Mote J, Gonzalez R, Abplanalp S, Zhang Y, Luckenbaugh J, et al. Smartphone sensing of social interactions in people with and without schizophrenia. J Psychiatr Res 2021 May;137:613-620. [doi: 10.1016/j.jpsychires.2020.11.002] [Medline: 33190842]
84. Wetzel B, Pryss R, Baumeister H, Edler J, Gonçalves AS, Cohrdes C. "How come you don't call me?" smartphone communication app usage as an indicator of loneliness and social well-being across the adult lifespan during the COVID-19 pandemic. Int J Environ Res Public Health 2021 Jun 08;18(12):6212 [FREE Full text] [doi: 10.3390/ijerph18126212] [Medline: 34201322]
85. Bellini G, Cipriano M, Comai S, De Angeli N, Gargano J, Gianella M, et al. Understanding social behaviour in a health-care facility from localization data: a case study. Sensors (Basel) 2021 Mar 18;21(6):2147 [FREE Full text] [doi: 10.3390/s21062147] [Medline: 33803913]
86. Petersen J, Austin D, Kaye JA, Pavel M, Hayes TL. Unobtrusive in-home detection of time spent out-of-home with applications to loneliness and physical activity. IEEE J Biomed Health Inform 2014 Sep;18(5):1590-1596. [doi: 10.1109/jbhi.2013.2294276]
87. Gao Y, Li H, Zhu T. Predicting subjective well-being by smartphone usage behaviors. In: Proceedings of the International Conference on Health Informatics. 2014 Presented at: Proceedings of the International Conference on Health Informatics; Mar 3-6, 2014; Loire Valley, France. [doi: 10.5220/0004800203170322]
88. Russell D, Peplau LA, Cutrona CE. The revised UCLA Loneliness Scale: concurrent and discriminant validity evidence. J Personality Soc Psychol 1980 Sep;39(3):472-480. [doi: 10.1037/0022-3514.39.3.472]
89. Mattick RP, Clarke J. Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. Behav Res Ther 1998 Apr;36(4):455-470. [doi: 10.1016/S0005-7967(97)10031-6]
90. Beaufort I, De Weert-Van Oene GH, Buwalda V, de Leeuw JR, Goudriaan A. The depression, anxiety and stress scale (DASS-21) as a screener for depression in substance use disorder inpatients: a pilot study. Eur Addict Res 2017 Dec 8;23(5):260-268 [FREE Full text] [doi: 10.1159/000485182] [Medline: 29224000]
91. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Personal Social Psychol 1988;54(6):1063-1070. [doi: 10.1037/0022-3514.54.6.1063]
92. Morahan-Martin J, Schumacher P. Loneliness and social uses of the internet. Comput Human Behav 2003 Nov;19(6):659-671. [doi: 10.1016/s0747-5632(03)00040-2]
93. Fokkema T, Knipscheer K. Escape loneliness by going digital: a quantitative and qualitative evaluation of a Dutch experiment in using ECT to overcome loneliness among older adults. Aging Ment Health 2007 Sep;11(5):496-504. [doi: 10.1080/13607860701366129] [Medline: 17882587]
94. Caplan SE. Relations among loneliness, social anxiety, and problematic internet use. Cyberpsychol Behav 2007 Apr;10(2):234-242. [doi: 10.1089/cpb.2006.9963] [Medline: 17474841]
95. Ceyhan AA, Ceyhan E. Loneliness, depression, and computer self-efficacy as predictors of problematic internet use. Cyberpsychol Behav 2008 Dec;11(6):699-701. [doi: 10.1089/cpb.2007.0255] [Medline: 19072150]

96. Morahan-Martin J. The relationship between loneliness and internet use and abuse. Cyberpsychol Behav 1999;2(5):431-439. [doi: 10.1089/cpb.1999.2.431] [Medline: 19178216]
97. Rossi A, Panzeri A, Pietrabissa G, Manzoni GM, Castelnuovo G, Mannarini S. The anxiety-buffer hypothesis in the time of COVID-19: when self-esteem protects from the impact of loneliness and fear on anxiety and depression. Front Psychol 2020 Nov 10;11:2177 [FREE Full text] [doi: 10.3389/fpsyg.2020.02177] [Medline: 33240140]
98. Þórarinsdóttir H, Kessing LV, Faurholt-Jepsen M. Smartphone-based self-assessment of stress in healthy adult individuals: a systematic review. J Med Internet Res 2017 Feb 13;19(2):e41 [FREE Full text] [doi: 10.2196/jmir.6397] [Medline: 28193600]
99. Batra S, Baker R, Wang T, Forma F, DiBiasi F, Peters-Strickland T. Digital health technology for use in patients with serious mental illness: a systematic review of the literature. Med Devices (Auckl) 2017 Oct;10:237-251 [FREE Full text] [doi: 10.2147/MDER.S144158] [Medline: 29042823]
100. Ko PT, Kientz JA, Choe EK, Kay M, Landis CA, Watson NF. Consumer sleep technologies: a review of the landscape. J Clin Sleep Med 2015 Dec 15;11(12):1455-1461 [FREE Full text] [doi: 10.5664/jcsm.5288] [Medline: 26156958]
101. Nguyen HH, Silva JN. Use of smartphone technology in cardiology. Trends Cardiovasc Med 2016 May;26(4):376-386. [doi: 10.1016/j.tcm.2015.11.002] [Medline: 26686266]
102. Kim BY, Lee J. Smart devices for older adults managing chronic disease: a scoping review. JMIR Mhealth Uhealth 2017 May 23;5(5):e69 [FREE Full text] [doi: 10.2196/mhealth.7141] [Medline: 28536089]
103. Faurholt-Jepsen M, Bauer M, Kessing LV. Smartphone-based objective monitoring in bipolar disorder: status and considerations. Int J Bipolar Disord 2018 Jan 23;6(1):6 [FREE Full text] [doi: 10.1186/s40345-017-0110-8] [Medline: 29359252]
104. Beutel ME, Klein EM, Brähler E, Reiner I, Jünger C, Michal M, et al. Loneliness in the general population: prevalence, determinants and relations to mental health. BMC Psychiatry 2017 Mar 20;17(1):97. [doi: 10.1186/s12888-017-1262-x] [Medline: 28320380]
105. Forkan AR, Branch P, Jayaraman PP, Ferretto A. An internet-of-things solution to assist independent living and social connectedness in elderly. Trans Soc Comput 2020 Jan 20;2(4):1-24. [doi: 10.1145/3363563]
106. Zhou J, Salvendy G. Human aspects of IT for the aged population. Design for the elderly and technology acceptance. In: Proceedings of the 5th International Conference, ITAP 2019, Held as Part of the 21st HCI International Conference, HCII 2019. 2019 Presented at: 5th International Conference, ITAP 2019, Held as Part of the 21st HCI International Conference, HCII 2019; Jul 26-31, 2019; Orlando, FL, USA. [doi: 10.1007/978-3-030-22012-9]
107. Uddin MZ, Khaksar W, Torresen J. Ambient sensors for elderly care and independent living: a survey. Sensors (Basel) 2018 Jun 25;18(7):2027 [FREE Full text] [doi: 10.3390/s18072027] [Medline: 29941804]
108. Macdonald SJ, Deacon L, Nixon J, Akintola A, Gillingham A, Kent J, et al. 'The invisible enemy': disability, loneliness and isolation. Disability Soc 2018 Sep 19;33(7):1138-1159. [doi: 10.1080/09687599.2018.1476224]
109. Emerson E, Fortune N, Llewellyn G, Stancliffe R. Loneliness, social support, social isolation and wellbeing among working age adults with and without disability: cross-sectional study. Disabil Health J 2021 Jan;14(1):100965 [FREE Full text] [doi: 10.1016/j.dhjo.2020.100965] [Medline: 32843311]
110. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". J Psychiatric Res 1975 Nov;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6]
111. Nait AA, Englebienne G, Kröse B. How lonely is your grandma? Detecting the visits to assisted living elderly from wireless sensor network data. In: Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication. 2013 Presented at: UbiComp '13: The 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing; Sep 8 - 12, 2013; Zurich Switzerland. [doi: 10.1145/2494091.2497283]
112. Martin S, Augusto JC, McCullagh P, Carswell W, Zheng H, Wang H, et al. Participatory research to design a novel telehealth system to support the night-time needs of people with dementia: NOCTURNAL. Int J Environ Res Public Health 2013 Dec 04;10(12):6764-6782 [FREE Full text] [doi: 10.3390/ijerph10126764] [Medline: 24304507]
113. Dennison L, Morrison L, Conway G, Yardley L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. J Med Internet Res 2013 Apr;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
114. Torous J, Friedman R, Keshavan M. Smartphone ownership and interest in mobile applications to monitor symptoms of mental health conditions. JMIR Mhealth Uhealth 2014 Jan;2(1):e2 [FREE Full text] [doi: 10.2196/mhealth.2994] [Medline: 25098314]
115. Harari GM, Müller SR, Mishra V, Wang R, Campbell AT, Rentfrow PJ, et al. An evaluation of students' interest in and compliance with self-tracking methods. Soc Psychol Personality Sci 2017 Jul 18;8(5):479-492. [doi: 10.1177/1948550617712033]
116. Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. PLoS One 2016 May 23;11(5):e0156164 [FREE Full text] [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]

Abbreviations

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
UCLA: University of California, Los Angeles

Edited by L Buis; submitted 02.11.21; peer-reviewed by J McMurray, I Mircheva, F Velayati; comments to author 28.12.21; revised version received 21.02.22; accepted 25.02.22; published 12.04.22.

Please cite as:
Qirtas MM, Zafeiridi E, Pesch D, White EB
Loneliness and Social Isolation Detection Using Passive Sensing Techniques: Scoping Review
JMIR Mhealth Uhealth 2022;10(4):e34638
URL: https://mhealth.jmir.org/2022/4/e34638
doi:10.2196/34638
PMID:35412465

Review

Accuracy and Precision of Energy Expenditure, Heart Rate, and Steps Measured by Combined-Sensing Fitbits Against Reference Measures: Systematic Review and Meta-analysis

Guillaume Chevance[1], PhD; Natalie M Golaszewski[2,3], PhD; Elizabeth Tipton[4], PhD; Eric B Hekler[2,3,5], PhD; Matthew Buman[6], PhD; Gregory J Welk[7], PhD; Kevin Patrick[3], PhD; Job G Godino[2,3,5,8], BSc, MSc, PhD

[1]Barcelona Institute for Global Health, ISGlobal, Barcelona, Spain

[2]Herbert Wertheim School of Public Health and Longevity Science, University of California, San Diego, La Jolla, CA, United States

[3]Center for Wireless & Population Health Systems, University of California, San Diego, La Jolla, CA, United States

[4]Department of Statistics, Northwestern University, Evanston, IL, United States

[5]Exercise and Physical Activity Resource Center, University of California, San Diego, La Jolla, CA, United States

[6]School of Nutrition & Health Promotion, Arizona State University, Phoenix, AZ, United States

[7]Department of Kinesiology, Iowa State University, Ames, IA, United States

[8]Laura Rodriguez Research Institute, Family Health Centers of San Diego, San Diego, CA, United States

Corresponding Author:
Job G Godino, BSc, MSc, PhD
Laura Rodriguez Research Institute
Family Health Centers of San Diego
823 Gateway Center Way
San Diego, CA, 92102
United States
Phone: 1 6195152344
Fax: 1 6195152344
Email: jobg@fhcsd.org

Abstract

Background: Although it is widely recognized that physical activity is an important determinant of health, assessing this complex behavior is a considerable challenge.

Objective: The purpose of this systematic review and meta-analysis is to examine, quantify, and report the current state of evidence for the validity of energy expenditure, heart rate, and steps measured by recent combined-sensing Fitbits.

Methods: We conducted a systematic review and Bland-Altman meta-analysis of validation studies of combined-sensing Fitbits against reference measures of energy expenditure, heart rate, and steps.

Results: A total of 52 studies were included in the systematic review. Among the 52 studies, 41 (79%) were included in the meta-analysis, representing 203 individual comparisons between Fitbit devices and a criterion measure (ie, n=117, 57.6% for heart rate; n=49, 24.1% for energy expenditure; and n=37, 18.2% for steps). Overall, most authors of the included studies concluded that recent Fitbit models underestimate heart rate, energy expenditure, and steps compared with criterion measures. These independent conclusions aligned with the results of the pooled meta-analyses showing an average underestimation of −2.99 beats per minute (*k* comparison=74), −2.77 kcal per minute (*k* comparison=29), and −3.11 steps per minute (*k* comparison=19), respectively, of the Fitbit compared with the criterion measure (results obtained after removing the high risk of bias studies; population limit of agreements for heart rate, energy expenditure, and steps: −23.99 to 18.01, −12.75 to 7.41, and −13.07 to 6.86, respectively).

Conclusions: Fitbit devices are likely to underestimate heart rate, energy expenditure, and steps. The estimation of these measurements varied by the quality of the study, age of the participants, type of activities, and the model of Fitbit. The qualitative conclusions of most studies aligned with the results of the meta-analysis. Although the expected level of accuracy might vary from one context to another, this underestimation can be acceptable, on average, for steps and heart rate. However, the measurement of energy expenditure may be inaccurate for some research purposes.

(JMIR Mhealth Uhealth 2022;10(4):e35626) doi:10.2196/35626

KEYWORDS

wearables; activity monitors; physical activity; validity; accelerometry

Introduction

Background

Although it is widely recognized that physical activity is an important determinant of health [1,2], assessing this complex behavior is a considerable challenge [3-5]. Tools for objective assessment of the frequency, intensity, and duration of physical activity in adults and children have largely been developed for short-term use within research or public health surveillance environments [6,7]. However, recent advances in microtechnology, data processing, wireless communication, and battery capacity have resulted in the proliferation of low-cost, noninvasive, wrist-worn devices with attractive designs that can be easily used by consumers to track their physical activity over long periods [8].

The latest generation of consumer-level activity monitors is typically multi-sensor devices that use *triaxial accelerometry* to measure movement and *photoplethysmography* to measure heart rate (ie, number of beats per minute [bpm]). Importantly, a combined-sensing approach to measuring physical activity may address many of the limitations of using either accelerometry or photoplethysmography alone [9,10]. The combination of these data streams through branched equation modeling or machine learned algorithms might result in a more accurate assessment of physical activity [11,12].

The expanding use of consumer-level activity monitors in population and clinical health research has led to an array of independent studies aimed at evaluating the validity of various metrics. No devices have received more attention than those manufactured by Fitbit (Fitbit Inc). From community-based health interventions that aim to motivate individuals to increase their physical activity level to interventions that aim to improve patient–health professional interactions, Fitbits are likely the most widely used [13,14]. Hence, a major concern for consumers and researchers alike is understanding the extent to which Fitbits provide accurate estimates of physical activity.

Several studies have evaluated the validity of different versions of Fitbits in estimating energy expenditure, intensity, heart rate, or steps, mostly in controlled laboratory settings [15] and a limited amount in free-living conditions [16]. Moreover, there have been 4 systematic reviews have been conducted to examine the accuracy of measures derived from consumer-level activity monitors in general [17-19] and from Fitbits specifically [20]. Taken together, these reviews conclude that Fitbit devices accurately measure steps and heart rate, whereas estimates of energy expenditure are less than optimal and tend to be underestimated. These reviews also spotted large variations around the estimates, highlighting potential sources of undetermined heterogeneity.

Although previous systematic reviews have been informative, several limitations exist within these reviews. First, 3 of the 4 systematic reviews [18-20] have compared Fitbits with questionable criterion measures, such as other wearable devices (ie, accelerometers), instead of ground truth or reference measures of energy expenditure [21], heart rate [22], or steps [23]. Second, all previous reviews have included older versions of the Fitbit that do not use photoplethysmography combined with accelerometry, which are (1) less likely to be used in future studies and (2) likely to result in more bias than the more recent Fitbits [11]. Third, there is yet to be a quantitative synthesis of the validity of recent Fitbits through a meta-analysis. Such meta-analytical work could notably help identify sources of heterogeneity in the validity of these devices for different outcomes and contexts of use.

Objective

The purpose of this systematic review and meta-analysis is to examine, quantify, and report on the current state of evidence for the analytical validity of energy expenditure, heart rate, and steps measured by recent combined-sensing Fitbits. On the basis of the existing literature, we expected some form of accuracy for the estimation of steps and heart rate and a lack of precision for energy expenditure. No hypotheses were formulated for the quantitative part of this study (ie, meta-analysis).

Methods

The protocol was registered with PROSPERO (International Prospective Register of Systematic Reviews; CRD42020161937) and is reported according to the PRISMA-P (Preferred Reporting Items for Systematic Review and Meta-Analyses Protocols) [24] guidelines (Multimedia Appendix 1). All study materials, including not only code and data but also the supplemental materials, are available on the *Open Science Framework* [25].

Search Strategy

A systematic review of the literature was conducted in 3 iterations to retrieve both published and unpublished studies [26]. The search was conducted using the PubMed and Embase databases from January 2015 (ie, commercialization of the first Fitbit device that included a heart rate monitor) to July 2021. The gray literature was also inspected through Open Grey (Multimedia Appendix 2). In the second iteration, studies were also sourced from previously published systematic reviews [17-20]. In the third iteration, reference lists within the studies included in the previous iterations were examined. Published conference abstracts were also included if sufficient detail was reported to assess study quality. In cases where information was missing, attempts were made to contact the authors. Study selection was performed by one coder (GC) and checked by an independent second coder (NMG). Any discrepancies were identified and resolved. No language restrictions were applied.

Criteria for Study Inclusion

Studies that simultaneously reported outcome data from a Fitbit device (energy expenditure, heart rate, or steps) and a valid criterion measure were considered. Only studies that evaluated

Fitbit devices that include a heart rate monitor (ie, *Charge HR 2015, Surge 2015, Blaze 2016, Charge 2 2016, Alta HR 2017, Ionic 2017, Versa 2018, Charge 3 2018, Inspire HR 2019, Versa 2 2019, Versa Lite Edition 2019, Charge 4 2020, Versa 3 2020, Sense 2020*, and *Inspire 2 2020*) were included. Valid criterion measures of energy expenditure included doubly labeled water or direct and indirect calorimetry; for heart rate, they included electrocardiograms, pulse oximeters, and specific chest-worn systems (eg, Polar), and for steps, direct observation was the only criterion (video recorded or not).

Data Extraction and Management

Information about the study characteristics (authors, year of publication, design, sample size, and number of observations for each outcome), population characteristics (age, health conditions, and BMI), descriptive statistics, type of Fitbit, and features of the criterion measures were extracted. Finally, given (1) the heterogeneity of the protocols to test the validity of the Fitbit, (2) the multiple statistical strategies used to perform the analyses (eg, Bland-Altman analyses vs analysis of variance), and (3) the lack of consensus in the interpretation of these statistical outcomes (ie, to infer whether a device is valid), we also decided to retrieve the explicit conclusion of the authors when judging the particular validity or a device.

For the meta-analysis, the effect sizes extracted were the mean bias (ie, accuracy) and variance or SD (ie, precision) in kilocalories per minute (kcal per minute), bpm, and difference of steps per minute (steps per minute) between the Fitbit and criterion measures of energy expenditure, heart rate, and steps, respectively. It is important to note that kcal and steps are not always reported as a function of time (ie, per minute). Some authors prefer the total amount of kcal or steps recorded during a specific task or an entire protocol. To make the comparisons between studies and interpretation of the results possible, we retrieved the time spent during each protocol task. We then converted the absolute number of kcal and steps to kcal and steps per minute by dividing the mean bias and SD reported by the duration of each specific task in minutes. For example, a mean bias of 20 (SD 10) kcal recorded over a 3-minute task was converted to 6 (SD 3) kcal per minute.

These outcomes were extracted directly from eligible studies when available or computed using other reported statistics (ie, means, SDs, and correlations). If needed, the authors were contacted and asked to provide the necessary information. Data were extracted and coded by one coder (GC) and checked by a second coder (NMG). Discrepancies were identified and resolved by rereferencing the articles and reaching a consensus with a third author (JGG).

Data Synthesis and Analyses

A specific meta-analytic framework was used for the analyses of agreement between the measures [27]. The main outcome of the Bland-Altman meta-analysis was the population limits of agreement between Fitbit devices and criterion measures of energy expenditure, heart rate, and steps. The population limits of agreement combine the bias of a test (ie, the average difference between the tested measure and a criterion measure) and the SD of these differences. The results from the individual studies were first converted into a standard format to conduct the meta-analysis, with bias captured as *Fitbit–criterion measure*. Outcomes were expressed in kcal per minute, bpm, and steps per minute for energy expenditure, heart rate, and steps, respectively.

The population limits of agreement were then computed to account for two sources of variation: the average within-study variation and the between-study variation. The computed population limits of agreement were typically wider (ie, more *conservative*) than those reported in other meta-analyses of Bland-Altman studies (for further explanations, refer to the study by Tipton and Shuster [27]). In this study, the pooled limits of agreement were calculated using $\delta \pm 2\sqrt{(\sigma^2+\tau^2)}$, where δ is the average bias across studies, σ^2 is the average within-study variation in differences, and τ is the SD of bias across studies (a larger τ indicates higher variations in bias between studies). Both δ and σ^2 were estimated using a weighted least squares model (similar to a random effects approach), and their SEs were estimated using robust variance estimation (RVE). RVE was used instead of model-based SEs as most of the studies included in our review used repeated measures designs without accounting for the correlation between measurements (ie, multilevel approach). The method of moments estimator was used for the τ parameter [28]. Measures of uncertainty were also included when interpreting the limits of agreement estimates by calculating the outer 95% CIs for pooled limits of agreement and adjusted repeated measurements, which were not properly adjusted for in individual studies [27]. Multiple effect sizes from the same study were also handled using the RVE method [29,30].

Planned Sensitivity and Subgroup Analyses

Subgroup meta-analyses were performed for the following variables: (1) characteristics of the participants, including the presence of health conditions and age (<65 years and >65 years); (2) type of Fitbit device; (3) type of activity (eg, resting and sedentary activities, ambulation, and cycling); (4) intensity (ie, differences in light and moderate to vigorous intensity activities); and (5) study quality (ie, see the following sections). The limits between light- and moderate-intensity physical activity for the intensity variable were defined according to the Compendium of Physical Activities. For example, walking >3 mph or 5 km/h and cycling >7 mph or 11 km/h, or 150 W, were considered moderate to vigorous physical activity. A complete description and justification of these analyses are provided in the registered protocol.

Quality Assessment (Risk of Bias)

A custom tool, developed based on a previous study using the COSMIN (Consensus-Based Standards for the Selection of Health Measurement Instruments) criteria [31], was used to assess study quality, including (1) sample size calculation justifying a reasonably large sample (N>50=1 point [32]), (2) peer reviewing (study peer reviewed=1 point), (3) appropriate placement of the device (device up to 3 finger widths above the wrist bone=1 point [33]), and (4) validation of only 1 device on the wrist (1 device at a time=1 point), thus providing a quality score between 1 (low) and 4 (high). Sensitivity analyses were

performed for the primary meta-analyses (ie, average energy expenditure, heart rate, and steps) based on the risk of bias by removing the *high risk of bias* studies (quality score ≤1) from the analyses and outliers. Subgroup analyses were also conducted according to the potential moderators identified previously and when at least four comparisons between the Fitbits and criterion measures were available.

All analyses were conducted using the R statistical program (version 4.1.2; R Foundation for Statistical Computing). The R code (adapted from the study by Tipton and Shuster [27]) and all the data used in the meta-analyses are available on the web [25].

Results

Systematic Review

A total of 52 studies were included in the systematic review (see Multimedia Appendix 3 for the study flowchart). Among the 52 studies, 41 (79%) were included in the meta-analyses, representing 203 individual comparisons between Fitbit devices and a criterion measure (ie, n=117, 57.6% for heart rate, n=49, 24.1% for energy expenditure, and n=37, 18.2% for steps; see study flowchart in Multimedia Appendix 3). The participants (n=1628) were mostly young (only 8/52, 15% of studies included participants aged >65 years), without chronic diseases (47/52, 90% of studies), and with a mean BMI of 24.9 kg/m^2 (range 21-34). Approximately 15% (8/52) of studies included participants with chronic conditions (ie, cardiac, respiratory, and Parkinson diseases and chronic pain). The included studies mostly tested the validity of the devices as part of formal and structured laboratory protocols (45/52, 87%; see the column *Protocol* in Table 1) instead of activities measured in free-living conditions.

Of the 52 studies, the Fitbit *Charge HR* was included in 27 (52%) studies, the *Surge* in 11 (21%) studies, the *Charge 2* in 10 (19%) studies, the *Blaze* and *Versa* in 3 (6%) studies each, and the *Ionic* and *Charge 3* in 1 (2%) study each. Of the 52 studies, Fitbits were compared with a criterion measure for heart rate in 32 (62%) studies, energy expenditure in 19 (37%) studies, and steps in 15 (29%) studies. According to our inclusion criteria, heart rate was mainly estimated using electrocardiograms (18/32, 56%) or Polar heart rate straps (14/32, 44%). Energy expenditure was estimated using indirect calorimetry in all studies except one, which used doubly labeled water. Steps were measured with video records for 57% (8/14) of studies and a manual hand counter for 43% (6/14) of studies.

Regarding the authors' study conclusions, 63% (20/32), 79% (15/19), and 27% (4/15) of studies concluded that the estimations provided by the Fitbit devices were not optimally valid compared with the reference standards for heart rate, energy expenditure, and steps, respectively. Most studies (18/32, 56%) explicitly reported an underestimation of the Fitbits compared with criterion measures for heart rate in their conclusion (only one of the studies explicitly reported an overestimation of heart rate; the remaining studies did not explicitly provide a conclusion about under- or overestimation). Similarly, a large number of studies (6/15, 40%) reported an underestimation of the Fitbits compared with criterion measures for steps (only one of the studies explicitly reported an overestimation of steps; the remaining studies did not explicitly provide a qualitative conclusion about under- or overestimation). Results were mixed for energy expenditure, with 12% (6/52) of studies explicitly reporting an underestimation of this outcome for the Fitbit, and 10% (5/52) reporting an overestimation (one of the studies indicated mixed findings related to the intensity and the remaining did not explicitly provide a conclusion about under- or overestimation). See Table 1 for a detailed description of each study included in the systematic review.

Table 1. Outcomes of the systematic review (N=52).

Study	Participants	Fitbit	Outcomes	Criterion measures	Protocol	Statistics	Authors' conclusion
Al-Kaisey et al [34]	Patients with cardiac conditions (N=12; observations=53,288)	Charge HR	HR[a]	ECG[b] (DigiTrakXT)	24-hour monitoring within a cardiology department (usual routine)	Correlations; multilevel Bland-Altman analyses	Underestimation, particularly pronounced at HR ranges >100 bpm[c]; accuracy judged as insufficient
Baek et al [35]	Healthy adults; mean age 24 years (N=15)	Charge 2	HR	ECG (Philips StressVue)	Two 20-minute walking sessions on a treadmill (1 conventional walking and 1 Nordic walking)	Bland-Altman analyses; Lin concordance correlation coefficients; mean relative difference; paired *t* test	Accuracy judged as adequate for conventional walking and inadequate during Nordic walking
Bai et al [36]	Healthy adults; aged 19 to 60 years (N=39)	Charge HR	HR; EE[d]; steps not used in the MA[e] (criterion measure=pedometer)	Polar heart rate chest strap; indirect calorimetry (Oxycon Mobile 5.0)	80-minute structured activity protocol (treadmill and free-living activities)	Bland-Altman analyses; MAPE[f]; equivalence testing	Accuracy judged as poor for EE but strong for HR
Bai et al [37]	Healthy adults; aged 18 to 59 years (N=48)	Charge 2	HR; steps not used in the MA (criterion measure=pedometer)	Polar heart rate chest strap	24-hour monitoring in a free-living setting (devices removed during the night)	Correlations; Bland-Altman analyses; MAPE; equivalence testing	Underestimation; accuracy judged as reasonable
Benedetto et al [38]	Healthy adults; aged 25 to 36 years (N=16; observations=9000)	Charge 2	HR	ECG (ProComp Infiniti T7500M)	Maximal 10-minute stationary bicycle test	Multilevel Bland-Altman analyses; ICC[g]	Underestimation; accuracy judged as poor
Boudreaux et al [39]	Healthy adults; aged 18 to 35 years (N=50)	Charge 2; Blaze	HR; EE not used in the MA (absolute value cannot be compiled)	ECG (Quinton 4500)	Structured activity protocol, including stationary cycling and resistance exercises (total time not provided)	MAPE; ICC; Bland-Altman analyses	Underestimation of HR judged as valid depending on the intensities and activities; accuracy of EE judged as inaccurate
Bunn et al [40]	Healthy adults; mean age 26 years (N=20)	Surge	Steps	Video recorded	10-minute walking and running bouts on a treadmill	MAPE; correlations; equivalence testing	Underestimation of steps above standards (MAPE<10%) for the walking bout and overestimation for the running bout; accuracy judged as poor for both intensities
Burton et al [41]	Healthy older adults; age >65 years (N=31)	Charge HR	Steps	Video recorded	2-minute walking tests; 2-week of measures in a free-living environment not used in the MA (criterion measure=accelerometer)	ICC; Bland-Altman analyses	Underestimation of steps; accuracy judged as good

Study	Participants	Fitbit	Outcomes	Criterion measures	Protocol	Statistics	Authors' conclusion
Cadmus-Bertram et al [42]	Healthy adults; aged 30 to 65 years (N=40)	Surge	HR	ECG (type not specified)	10-minute treadmill exercise at 65% of the maximum HR	Multilevel Bland-Altman analyses	Accurate agreement at rest; poor agreement when participant exercised at 65% of their maximum HR; overall accuracy judged as insufficient
Chow et al[h] [43]	Healthy adults; mean age 24 years (N=31)	Charge HR	Steps	Manual hand counter	3-minute treadmill exercise at varying speeds	ANOVA[i]	Underestimation of steps at slowest speeds; accuracy improved at faster speeds; no clear conclusion about the overall accuracy of the device
Chowdhury et al [44]	Healthy adults; aged 18 to 50 years (N=30)	Charge HR	EE	Indirect calorimetry (COSMED K4b2)	Simulated activities of daily living and structured exercise in laboratory conditions (64-minute in total); 24-hour period in free-living conditions not used in the MA (criterion measure=accelerometers and armband device)	Bland-Altman analyses; mean signed error tests; MAE[j] tests; correlations; ANOVA; equivalence testing	Underestimation of EE in the 2 conditions; not as consistent as research-grade devices
Claes et al [45]	Healthy adults; aged 18 to 40 years (N=18)	Charge HR	EE; steps	Indirect calorimetry (Jaeger Oxycon Mobile); video recorded	50-minute protocol on a treadmill at various intensities	Paired sample *t* tests; Wilcoxon signed ranks tests; Bland-Altman analyses	Estimation of the 2 outcomes judged as accurate
Herkert et al [46]	Patients with cardiac conditions (N=19)	Charge 2	EE	Indirect calorimetry (Jaeger Oxycon Mobile)	Low- to moderate-intensity walking and cycling activities (protocol duration not provided)	Bland-Altman analyses; ICC	Accuracy judged as poor
Düking et al [47]	Healthy adults; mean age 26 years (N=25)	Versa	HR; EE	Polar HR chest strap; indirect calorimetry (Metamax 3B, CORTEX Biophysik GmbH)	5 minutes of sitting, walking, and running at different velocities and intermittent sprints during 3 minutes performed on a treadmill	Standardized mean bias; standardized typical error of the estimate; coefficient of variation; Pearson correlation	HR should be interpreted with caution because of the high error rate, and the Fitbit should not be used to monitor EE

Study	Participants	Fitbit	Outcomes	Criterion measures	Protocol	Statistics	Authors' conclusion
Dooley et al [15]	Students; aged 18 to 38 years (N=62)	Charge HR	HR; EE	Polar HR chest strap; indirect calorimetry (Parvo Medics TrueOne 2400)	40-minute treadmill protocol performed at various intensities	ANOVA; Bland-Altman analyses; MAPE	Overestimation of HR during light-intensity activities and overestimation of EE during light and moderate intensities; accuracy judged as reasonably accurate to estimate HR but not accurate for EE
Etiwy et al [48]	Patients with cardiac conditions; mean age 62 years (N=80)	Blaze	HR	ECG (type not specified)	15-minute treadmill protocol performed at various intensities	MAPE; Bland-Altman analyses; correlations; mixed model analyses of variance	Underestimation of HR; accuracy judged as probably insufficient among patients with cardiac conditions
Falgoust et al[h] [49]	Healthy adults; aged 23 to 54 years (N=30)	Charge HR; Surge	Steps	Manual hand counter	2×2 laps on a track at a self-selected walking speed	ANOVA; correlations	Underestimation of steps, more pronounced for the Fitbit Surge than the Charge HR; accuracy judged as insufficient for research purpose
Fokkema et al [50]	Healthy adults; mean age 32 years (N=31)	Charge HR	Steps	Manual hand counter	Two 30-minute treadmill walking bouts at 3 different walking speeds	ICC; MAPE; paired sample *t* tests; Wilcoxon signed-rank tests	Accuracy decreased as walking speed increased; accuracy was judged as not valid for high walking speeds but acceptable for lower walking speeds
Gaynor et al [51]	Patients with respiratory conditions; mean age 34 years (N=15)	Charge HR	HR	ECG (type not specified)	One 15-minute session of continuous cycling on an ergometer and one 15-minute session of interval cycling	Bland-Altman analyses	Underestimation, particularly pronounced during continuous exercise compared with interval training; authors recommended to not use a Fitbit Charge HR for assessing HR during exercise in adults with cystic fibrosis

Study	Participants	Fitbit	Outcomes	Criterion measures	Protocol	Statistics	Authors' conclusion
Gillinov et al [52]	Healthy adults; mean age 38 years (N=50; observations=3985)	Blaze	HR	ECG (type not specified)	24-minute structured exercise protocols on a treadmill, ergometer, and elliptical trainer	Correlations; MAPE; Bland-Altman analyses; mixed model analyses of variance	Accuracy varies with the activities but, overall, judged mostly inaccurate
Gorny et al [16]	Healthy adults; mean age 25 years (N=10; observations=2769)	Charge HR	HR	Polar HR chest strap	3 to 6 hours of normal daily living activities	ICC; Multilevel Bland-Altman analyses	Underestimation, particularly pronounced for higher intensity activities; accuracy inconclusive
Jagim et al [53]	Healthy adults; mean age 24 years (N=20)	Versa	HR; EE	ECG (12-lead CareCenter MD ECG); indirect calorimetry (TrueMax 2400 Metabolic Measurement System, Parvo-Medics)	12-minute graded exercise protocol at speeds of 4.8 km/hour, 7.2 km/hour, 9.6 km/hour, and 12.1 km/hour on a motorized treadmill	Pearson correlation; ANOVAs; MAPE; constant error; Bland-Altman analyses	Underestimation of HR and overestimation of EE; no clear conclusion about the overall accuracy of the device
Jo et al [54]	Healthy adults; mean age 24 years (N=24; observations=87,340)	Charge HR	HR	ECG (Cosmed C12x)	77-minute protocol comprising various activities (treadmill, ergometer, and resistance) performed at two intensities (light and moderate to vigorous)	Correlations; multilevel Bland-Altman analyses; MAPE	Underestimation at the higher ends of the mean HR spectrum; failed to satisfy validity criteria
Lai et al [55]	Patients with Parkinson disease; mean age 64 years (N=31)	Charge 2	Steps	Manual hand counter	6-minute bouts of overground and treadmill walking at a comfortable speed	ICCs; Bland-Altman analyses; MPE[k]	Accurate and precise for overground walking only
Lamont et al[h] [56]	Patients with Parkinson disease; mean age 69 years (N=33)	Charge HR	HR; steps not used in the MA (criterion measure=accelerometer)	Polar HR chest strap	Six 2-minute walking bouts at various intensities on an indoor track	MAPE; Bland-Altman analyses; paired sample *t* tests	Weakly associated with increases in HR; no clear conclusion about the overall accuracy of the device
Lee et al [57]	Students; mean age 27 years (N=10)	Charge HR	HR	Polar HR chest strap	8-hour continuous monitoring during normal daily activities	Correlations; MAPE; Multilevel analyses of variance	Measurement judged as inaccurate
Modave et al[h] [58]	Healthy adults in three age groups: 18 to 39 years, 40 to 64 years, 65 to 84 years (N=60)	Surge	Steps	Manual hand counter	Two separate 1000-step walks on a treadmill at a self-selected speed	Multilevel analyses of variance	Underestimation of steps across all age groups
Montes et al [59]	Healthy adults; mean age 25 years (N=40)	Surge	Steps	Manual hand counter	5-minute walking and running free motion and treadmill	MAPE; Bland-Altman analyses; Pearson correlation; ICC	Underestimation of steps for all activities, with walking activities being higher than the running; valid for all conditions except treadmill walking

Study	Participants	Fitbit	Outcomes	Criterion measures	Protocol	Statistics	Authors' conclusion
Montoye et al [60]	Healthy adults; mean age 24 years (N=32)	Charge HR	HR; EE; steps not used in the MA (criterion measure=pedometer)	Pulse oximeter (Nonin PureSAT); indirect calorimetry (Parvo TrueOne 2400)	90-minute structured protocol performed at various intensities in laboratory condition and on a 200 m indoor track	ANOVA; paired sample *t* tests; MAPEs; Bland-Altman analyses	Underestimation of HRs for higher intensity activities and poor estimation of EE
Morris et al [61]	Healthy adults; mean age 29 years (N=47)	Charge HR	EE	Indirect calorimetry (Cosmed K4b2)	15-minute high-intensity workout	ICC; ANOVA; MAPE	Significant underestimation of EE; judged as inaccurate
Muggeridge et al [62]	Healthy adults; mean age 40 years (N=20; *k*=35,639)	Charge 3	HR	Polar HR chest strap	Visit 1: 15-minute sedentary activities, 10-minute cycling on a bicycle ergometer, and incremental exercise test to exhaustion on a motorized treadmill; visit 2: four 15-second maximal sprints on a cycle ergometer and four 30 m to 50 m sprints on a treadmill	Multilevel Bland-Altman analyses; MAPE; Pearson correlation	Accuracy was generally poor, notably, during cycling exercises; underestimation of HR
Nelson and Allen [63]	Healthy adults; mean age 29 years (N=1; *k*=102,740)	Charge 2	HR	ECG (Vrije Universiteit Ambulatory Monitoring System)	24 hours of daily living monitoring	Multilevel Bland-Altman analyses; MAPE; CCC[l]	Slight underestimation; judged as acceptable
Nuss et al [64]	Healthy adults; mean age 24 years (N=20)	Charge 2	EE	Indirect calorimetry (Parvo Medics TrueOne 2400)	Bruce treadmill protocol (maximal)	CCC; MAPE; paired sample *t* tests	Significant underestimation judged as inaccurate
Pasadyn et al[h] [65]	Healthy adults; mean age 29 years (N=50)	Ionic	HR	ECG (Quinton Q-tel RMS telemetry system)	12-minute treadmill protocol performed at various intensities	CCC; Bland-Altman analyses; mixed model ANOVA	Moderate to high level of accuracy
Powierza et al [66]	Healthy adults; age range 18 to 26 years (N=22)	Charge HR	HR	ECG (MP150, BioPac Systems)	Buffalo Concussion Treadmill Test (maximal)	ICC; multilevel Bland-Altman analyses; MAPE	Small underestimation judged as not accurate for monitoring HR within a narrow range
Pribyslavska et al[h] [67]	Healthy adults; mean age 26 years (N=34)	Surge	EE	Indirect calorimetry (Oxycon Mobile)	Two 2-minute bouts on an ergometer and treadmill at different intensities	PE[m]; MAPE	Underestimation; accuracy judged as reasonable
Reddy et al [68]	Healthy adults; mean age 28 years (N=20)	Charge 2	HR; EE	Polar HR chest strap; indirect calorimetry (Cosmed K4b2 or Cosmed K5)	Maximal oxygen uptake test, resistance exercises, interval training (27 minutes), and free-living activities (28 minutes)	MAPE; Bland-Altman analyses; correlations	Underestimation; accuracy judged as reasonable
Salazar et al [69]	Healthy adults; mean age 22 years (N=35)	Charge 2	HR	Polar HR strap	12-minute treadmill protocol at different intensities	ANOVA; correlations	Underestimation; accuracy judged as adequate

Study	Participants	Fitbit	Outcomes	Criterion measures	Protocol	Statistics	Authors' conclusion
Shcherbina et al[h] [70]	Healthy adults; mean age 38 years (N=60)	Surge	HR; EE	ECG (type not specified); indirect calorimetry (Quark CPET, COSMED)	38-minute treadmill and ergometer protocol performed at various intensities	PE; ANOVA; principal component analysis; correlations; Bland-Altman analyses	Measure of HR judged adequate but poor for EE
Siddall et al [71]	Military officer trainees; mean age 23 years (N=20)	Surge	EE	Doubly labeled water	10 days of military training	Correlations; Bland-Altman analyses	Underestimation judged as insufficiently accurate
Sjöberg et al [72]	Adults with chronic pain; mean age 44 years (N=41)	Versa	EE; HR	Indirect calorimetry (Jaeger Oxycon Pro); Polar HR strap	Treadmill walking at three speeds (3.0 km/hour, 4.5 km/hour, and 6.0 km/hour) in the laboratory setting	ICC; ANOVA; Bland-Altman; MAPE	Overestimation of EE; accuracy judged as poor; good agreement for HR that tends to decrease with speed
Stahl et al [73]	Healthy adults; age range 19 to 45 years (N=50; observations=1781)	Charge HR	HR	Polar HR strap	30-minute treadmill protocol performed at different intensities	Correlations; multilevel Bland-Altman analyses; MAPE; ANOVA; equivalence testing	Small underestimation; accuracy judged as adequate
Tam and Cheung [74]	Healthy adults; mean age 32 years (N=30)	Charge HR	Steps	Video recorded	25-minute treadmill protocol performed at different intensities	Paired sample *t* tests; Bland-Altman analyses; correlations; MAPE	Estimation judged as accurate
Tedesco et al [75]	Older adults; mean age 69 years (N=18)	Charge 2	Steps; HR	Video recorded; Polar HR chest strap	3-hour structured protocol involving walking on a treadmill, simulated household, and sedentary activities	Mean bias; MPE; MAPE; MAD[n]; MAE; RMSE[o]; ICC; paired sample *t* tests; Wilcoxon signed-rank test; Bland-Altman analyses	Underestimation of heart rate; deficits in accuracy
Thiebaud et al [76]	Healthy adults; mean age 22 years (N=22)	Surge	HR; EE	ECG (Quinton Q-Stress, version 4.5); indirect calorimetry (Trueone 2400, Parvomedics)	15-minute treadmill protocol performed at various intensities	Correlations; limits of agreement; MAPE; equivalence testing	Underestimation of HR judged as acceptable; overestimation of EE at each speed and judged as insufficiently accurate
Thomson et al [77]	Healthy adults; mean age 24 years (N=30)	Charge HR	HR	ECG (Q-Stress, Mortara)	Bruce treadmill protocol (maximal)	Equivalence testing; CCC; Bland-Altman analyses	Underestimation increasing with intensity; Overall accuracy judged as insufficient

Study	Participants	Fitbit	Outcomes	Criterion measures	Protocol	Statistics	Authors' conclusion
Tophøj et al[h] [78]	Healthy students; mean age 26 years (N=20)	Surge; Charge HR	Steps	Manual hand counter	800 steps performed on a treadmill	MAPE; ICC; Bland-Altman analyses	Accurate estimation for the Fitbit Surge at higher walking speeds and inaccurate estimations at lower speeds; the Fitbit Charge HR was judged as insufficiently accurate
Wahl et al[h] [79]	Healthy sport students; mean age 25 years (N=20)	Charge HR	Step count; EE	Manual hand counter; indirect calorimetry (Metamax 3B, CORTEX Biophysik GmbH)	55-minute treadmill protocol at constant and intermittent velocities; outdoor exercise not included in the MA	MAPE; ICC; TE^{p}; Bland-Altman analyses	Acceptable level of validity for steps; inaccurate estimation of EE, with overestimation of EE for slower velocities and underestimation of EE for higher velocities
Wallen et al [80]	Healthy participants; mean age 24 years (N=22)	Charge HR	HR; EE; steps	ECG (CASE, GE Healthcare); indirect calorimetry (MetaMax 3B, Cortex); video recorded	58-minute treadmill and ergometer protocol performed at various intensities	Correlations; Bland-Altman analyses	Accurate measure of HR; overestimation of EE, judged as inaccurate; no clear conclusion is proposed for steps
Wang et al[h] [81]	Healthy participants; mean age 37 years (N=50; observations=1773)	Charge HR	HR	Polar HR chest strap	18-minute treadmill protocol at various intensities	CCC; Wilcoxon signed-rank; Bland-Altman analyses	Adequate estimation of HR at low intensities, suboptimal accuracy during moderate exercise, and underestimated during vigorous exercise; judgment deemed as inaccurate
Xie et al[h] [82]	Healthy participants; aged 19 to 27 years (N=44)	Surge	Steps; EE; HR not used in the MA (criterion measure=manual estimation)	Indirect calorimetry (Cosmed K4b2); video recorded	Walking, running, and cycling on a 400 m standard track	MAPE; correlations; paired sample *t* tests	High accuracy of measure for steps; inadequate accuracy of EE

Study	Participants	Fitbit	Outcomes	Criterion measures	Protocol	Statistics	Authors' conclusion
Zhang et al [83]	Healthy students; mean age 20 years (N=30)	Charge HR	EE	Indirect calorimetry (TrueOne 2400, Parvo Medics Inc)	6 structured 10-minute exercise bouts on a treadmill at various intensities	Paired sample *t* tests; equivalence testing; correlations; MAPE; Bland-Altman analyses	Estimation of EE judged as adequate during treadmill running

[a]HR: heart rate.
[b]ECG: electrocardiogram.
[c]bpm: beats per minute.
[d]EE: energy expenditure.
[e]MA: meta-analysis.
[f]MAPE: mean absolute percentage error.
[g]ICC: intraclass correlation coefficient.
[h]Studies not included in the meta-analysis.
[i]ANOVA: analysis of variance.
[j]MAE: mean absolute error.
[k]MPE: mean percentage error.
[l]CCC: concordance correlation coefficient.
[m]PE: percentage error.
[n]MAD: median absolute deviation.
[o]RMSE: root mean square error.
[p]TE: typical error.

Meta-analyses

Table 2 presents the results of the main and sensitivity analyses after removing studies with a high risk of bias (ie, low quality). Regarding heart rate, the pooled estimate of the mean bias between Fitbit devices and criterion measures was −3.39 bpm (*k* comparison=117), indicating an underestimation of the Fitbits compared with criterion measures. The range in population limits of agreement was large, resulting in the 2 methods differing from −24 bpm to 18 bpm across all studies. Underestimation slightly improved when removing low-quality studies (*k* comparison=74) from −3.39 bpm for the main analysis to −2.99 bpm (however, heterogeneity remained similar).

Regarding steps, the mean bias between Fitbit devices and criterion measures was −1.47 steps per minute, indicating an underestimation of the Fitbits compared with the criterion measures (*k* comparison=37). The population limit of agreement was large, ranging from −15 steps per minute to 12 steps per minute across all studies. These differences were more pronounced after removing studies with a low-quality score but with a lower heterogeneity (*k* comparison=19): pooled estimate of −3.11 steps per minute ranged between −13 steps per minute and 7 steps per minute.

Figure 1 displays the results (main meta-analyses and sensitivity analyses) as a forest plot. Figure 1 highlights the particularly high heterogeneity for heart rate compared with energy expenditure and steps. This heterogeneity is addressed in the following section using a series of subgroup analyses.

Table 2. Results of the main meta-analysis.

Analyses	*k* comparisons[a]	Bias[b], mean (SD[c])	τ[d]	LoA[e,f]	95% CI[g]
Main analyses					
HR[h] (bpm[i])	117	−3.39 (9.91)	11.35	−24.32 to 17.53	−26.36 to 19.58
EE[j] (kcal per minute)	49	0.19 (2.53)	0.99	−5.32 to 5.70	−7.23 to 7.61
Steps (per minute)	37	−1.47 (6.30)	6.50	−15.07 to 12.13	−20.55 to 17.61
Low-quality studies removed					
HR (bpm)	74	−2.99 (9.43)	21.42	−23.99 to 18.01	−27.68 to 21.71
EE (kcal per minute)	29	−2.77 (4.12)	8.40	−12.75 to 7.41	−15.28 to 9.95
Steps (per minute)	19	−3.11 (4.32)	6.17	−13.07 to 6.86	−17.27 to 11.06
Outliers removed					
HR (bpm)	116	−3.34 (9.79)	11.35	−24.06 to 17.37	−26.09 to 19.40
EE (kcal per minute)	48	0.19 (2.38)	0.98	−4.96 to 5.38	−6.68 to 7.06
Steps (per minute)	36	−1.02 (6.07)	6.17	−14.15 to 12.11	−19.49 to 17.46

[a]*k* comparisons is the number of comparisons between the Fitbits and criterion measures available within studies.

[b]Bias is the pooled estimate of mean differences calculated as Fitbit–criterion measures.

[c]SD is the pooled SD of differences.

[d]τ is the variation in bias between studies.

[e]LoA: limits of agreement.

[f]Lower 95% limit of agreement calculated from pooled estimates of bias and SD of differences with robust variance estimation and upper 95% limit of agreement calculated from pooled estimates of bias and SD of differences with robust variance estimation.

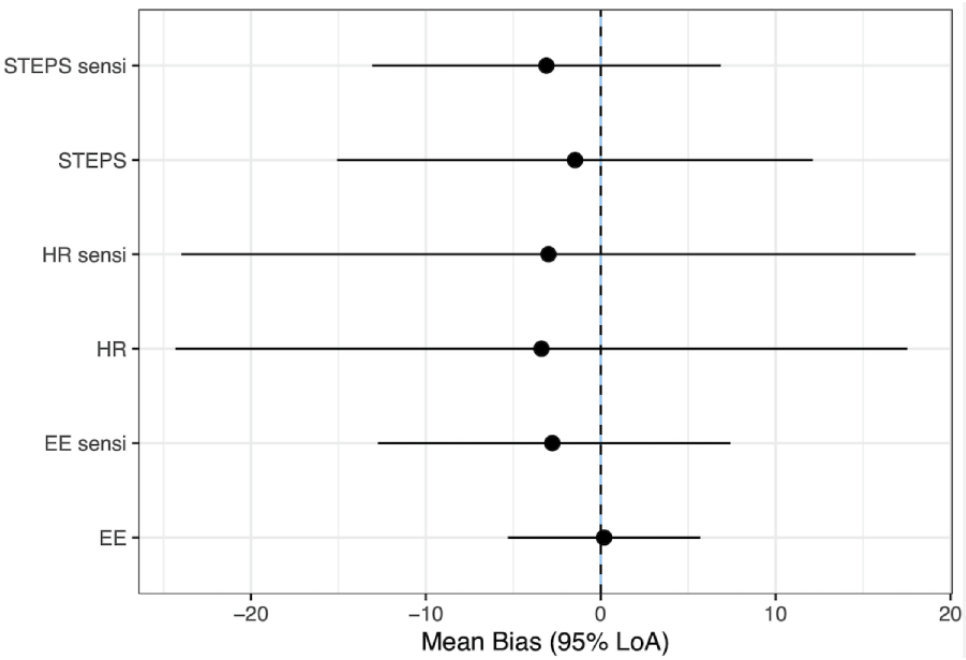
[g]Outer confidence bound for lower 95% limit of agreement and outer confidence bound for the upper 95% limit of agreement.

[h]HR: heart rate.

[i]bpm: beats per minute.

[j]EE: energy expenditure. Regarding energy expenditure, the mean bias between Fitbits and criterion measures was 0.19 kcal per minute, and the range in population limits of agreement was large, between −5 kcal per minute and 6 kcal per minute across participants (*k* comparison=49). This result is somewhat inconsistent with the meta-analysis, excluding low-quality studies (*k* comparison=29), which indicated an underestimation of the Fitbit of −2.77 kcal per minute (population limits of agreement comprise between −13 kcal per minute and 7 kcal per minute).

Figure 1. Forest plots for the main and sensitivity analyses. EE: energy expenditure; HR: heart rate; LoA: limits of agreement.



Subgroup Analyses

A range of subgroup analyses is presented in Tables S1, S2, and S3 in Multimedia Appendix 4 and can be visualized altogether in Figure 2. Overall, subgroup analyses by population characteristics, intensities, and types of activities, as well as Fitbits' models, were consistent with the main findings (ie, showing an underestimation of the Fitbits compared with criterion measures in most cases).

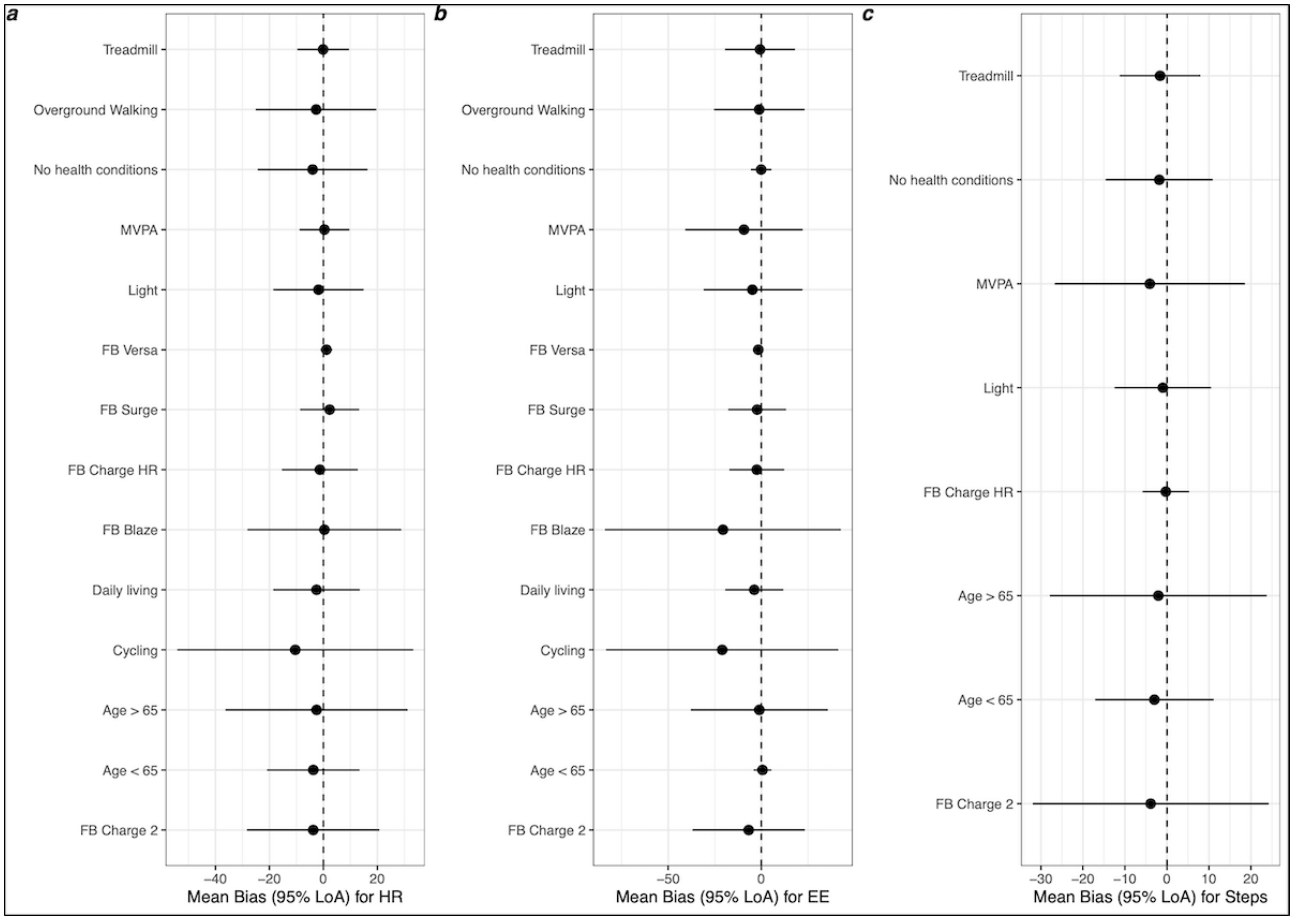
Compared with young and middle-aged adults, the results indicated a relatively similar mean bias in the 2 age groups, as well as in the subgroup of participants without health conditions (these results should be considered with caution, given the disproportionately lower number of studies conducted in older adults; *k* comparisons were between 6 and 26). Heterogeneity in these effects (ie, 95% limits of agreement) was systematically lower in younger than in older adults and lower in participants without health conditions, particularly for energy expenditure (Table S1 in Multimedia Appendix 4; Figure 2).

The results of the subgroup meta-analyses for different intensities and types of activities (Table S2 in Multimedia Appendix 4) clearly show a more pronounced underestimation of heart rate and energy expenditure for cycling activities compared with daily living and treadmill activities as well as overground walking. Performance of the device was better (lower heterogeneity) for treadmills than for overground walking. For energy expenditure and steps, the underestimation, and heterogeneity of these effects, were larger for moderate to vigorous intensity activities than for light-intensity activities. Opposite results were observed for heart rate, with more accurate measurements (ie, smaller bias and lower heterogeneity) at moderate to vigorous intensity activities compared with light-intensity activities.

The results of the subgroup meta-analyses by type of device and considering the number of *k* comparisons available by device show that the Fitbit Charge HR presents better performance than other models, notably in comparison with the Fitbit Charge 2 that has been tested in a comparable number of studies (Table S3 in Multimedia Appendix 4). Performance of the Fitbit Charge HR was particularly good for steps, with a mean bias of −0.27 steps per minute ranging between −6 steps per minute and 5 steps per minute. Interestingly, the Fitbit Versa was particularly precise compared with other models (Figure 2); however, this result should be confirmed on the basis of more future validation studies for this specific device.

Figure 2 displays the results of the subgroup meta-analyses for heart rate (Figure 2A), energy expenditure (Figure 2B), and steps (Figure 2C).

Figure 2. Forest plots for the subgroup analyses. EE: energy expenditure; FB: Fitbit; HR: heart rate; LoA: limits of agreement; MVPA: moderate to vigorous physical activity.



Discussion

Principal Findings

The results of this systematic review and meta-analysis showed that Fitbit devices are likely to underestimate heart rate, energy expenditure, and steps. This work adds to the current state of evidence for the analytical validity of heart rate, energy expenditure, and steps measured by recent combined-sensing Fitbits compared with criterion measures, many of which are considered *gold standards* or widely used reference standards. This is also the first review to include meta-analyses of Bland-Altman results evaluating the validity of measures of heart rate, energy expenditure, and steps for these devices. Thus, it offers actionable quantitative information to appreciate device validity.

Overall, our systematic review revealed that most authors of the included studies concluded that Fitbits underestimated heart rate, energy expenditure, and steps compared with criterion measures (Table 1). These independent (qualitative) conclusions aligned with the results of our meta-analysis, even in sensitivity and subgroup analyses that considered various aspects of study quality. The fact that results from the authors' qualitative conclusion (obtained via our systematic review) and this meta-analysis aligned is important, given the heterogeneity of study designs and statistical procedures used in the literature. The underestimation of activity intensity appears consistent with previous systematic reviews, including different brands of activity monitors, older Fitbits, and/or other criterion measures than those considered in this study (see the study by O'Driscoll et al [17] for energy expenditure, the study by Evenson et al [18] for steps, and the studies by Fuller et al [19] and Feehan et al [20] for the 3 outcomes).

However, precisely interpreting the magnitude of this underestimation remains a challenge, as there is little consensus in the literature regarding what constitutes an acceptable magnitude of bias or error. As observed in this systematic review, the interpretations and conclusions from the authors of the included studies were highly variable from one study to another (ie, a result deemed acceptable in one study can be judged as poor in another). Excluding low-quality studies, our pooled estimates indicated that Fitbits underestimate by approximately 3 bpm, 3 steps per minute, and 3 kcal per minute compared with the respective criterion measures. The implications of these differences depend on the nature of the comparisons and on the application. For heart rate, an underestimation of 3 bpm may be an acceptable difference, as the Association for the Advancement of Medical Instrumentation has defined the accuracy of cardiac monitors, heart rate meters, and alarms as a readout error of no greater than +5 and –5 bpm [84]. A similar interpretation can be provided for steps. Assuming that the average 3 steps per minute bias is linear over time and intensities, a 1-hour walk would result in an average underestimation of 180 steps (3 steps × 60 minutes). At a pace of 100 steps per minute (which corresponds to a moderate-intensity walk for the general population [85]), the Fitbit would indicate 5820 steps instead of 6000, which might be judged as a relatively small underestimation of 3% (ie, 5820×100/6000). However, a mean bias of 3 kcal per minute might be met with greater concern. Applying a similar logic as for the steps, after 1 hour of a specific activity, the Fitbit would detect an average of −180 kcal per minute. This is the estimated difference between a 1-hour walk at 3.5 mph to 4.5 mph for a 154 lbs (70 kg) person (respectively 280 kcal per hour and 460 kcal per hour [86]), representing an underestimation of approximately 40% (ie, 280×100/460).

The approximately 3 units of underestimation referred to above may vary largely within participants as well between studies and contexts (as indicated by the large pooled limit of agreement and their CIs, as well as the variation τ in bias between studies). According to our subgroup analyses, this heterogeneity is higher (1) in older adults than in younger adults and adults without chronic health conditions, (2) for cycling activities than for other activities, and (3) for the Fitbit Charge 2 than for the Fitbit Charge HR (ie, the 2 devices that received the most attention in the literature). Noticeable results also include reduced heterogeneity (ie, better validity) for energy expenditure in younger adults, heart rate for moderate to vigorous intensities, and Fitbit Charge HR for steps. Other potential differences must be taken with caution, given the number of comparisons (k) available per subgroup analysis. Replicating these subgroup analyses with an individual participant meta-analysis approach (ie, meta-analyzing each participant's estimates instead of the studies' pooled estimates) would constitute an interesting next step to even more precisely quantify the heterogeneity in these effects. However, this would require a greater number of open-access data sets from researchers in this specific field, which is not the case for now.

This study also highlights the need for ongoing high-quality validation research that uses a greater level of protocol standardization, particularly in regard to the assessment tasks, criterion measures, and reported analyses, following, for instance, the ones recommended in the study by Welk et al [87]. Consensus-building efforts that are focused on methodological rigor among researchers in this field are warranted, as are efforts to establish acceptable ranges of accuracy for the metrics of interest. The adoption of common practices for validation studies would facilitate the conduct of robust meta-analyses with comparable metrics and outcomes. In addition, protocols that systematically isolate a wide range of suggested sources of bias (eg, device movement, arm hair, sweat, skin thickness, skin tone, and adiposity) that may affect the underlying technologies in most wrist-worn multi-sensor devices (ie, accelerometry and photoplethysmography) are needed. Finally, as previously mentioned, the adoption of open science practices, notably data sharing, would greatly facilitate future meta-analyses of individual studies.

Limitations and Perspectives

This systematic review and meta-analysis is not without limitations. First, we restricted our synthesis to studies of adults, as although the number of studies that include children is growing, there remains a dearth of high-quality studies in this area. Additional research across the age span is needed to close the gap in our understanding of how well the Fitbits measure physical activity in young individuals and older adults. Second,

many different statistical strategies and related effect sizes are used to estimate the validity of these devices [37]. Researchers have used, separately or in combination, analysis of variance, correlations (eg, intraclass coefficient correlation), and measures of agreement (eg, Bradley-Blackwood test, Bland-Altman analyses, and mean absolute percentage error). At present, there is no specific framework for meta-analyzing statistics, such as mean absolute percentage error, although it is a preferred metric for understanding validity [37]. Thus, the meta-analysis was restricted to the mean bias and SD from the Bland-Altman analyses. Third, the field of physical activity measurement has yet to establish the magnitude of bias from consumer-level activity monitors that is acceptable or problematic. These classifications are likely contingent on the context in which the devices are used. For example, if one is using a consumer-level activity monitor for self-monitoring within a physical activity promotion intervention, a modest underestimation might not have a large negative impact on the research. However, underestimation within epidemiological surveillance efforts is less than ideal. A consensus regarding the magnitude of error that is either acceptable or unacceptable within a given research context would allow for improved interpretation of the results of validation efforts. Finally, to make comparisons between studies, we retrieved the time spent during each protocol task and converted the absolute number of kcal and steps to kcal per minute and steps per minute. This analytical strategy is not without limitations, notably for energy expenditure. This assumes that energy expenditure is linear over time and over a protocol, which may not be the case.

Conclusions

Compared with reference standards, recent Fitbit devices are likely to underestimate heart rate, energy expenditure, and steps by an average of three units per minute (ie, steps, bpm, and kcal). Although the expected level of accuracy might vary from one context to another, this underestimation can be acceptable, on average, for steps and heart rate. However, the measurement of energy expenditure may be too inaccurate for some research purposes. The estimation of these measurements varied slightly by the quality of the study, age of the participants, type of activities, and model of Fitbit. Overall, devices were more accurate in younger adults, for treadmills activities (notably, compared with cycling), and for the Fitbit Charge HR (notably, for steps).

Acknowledgments

GC was supported by the Spanish Ministry of Science and Innovation and State Research Agency through the *Centro de Excelencia Severo Ochoa 2019-2023* program (CEX2018-000806-S) and the Generalitat de Catalunya through the CERCA program. NMG was supported by a grant from the National Institute of Health's National Institute on Aging (T32AG058529).

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-P (Preferred Reporting Items for Systematic Review and Meta-Analyses Protocols) checklist.
[DOCX File , 31 KB - mhealth_v10i4e35626_app1.docx]

Multimedia Appendix 2
Search terms.
[DOCX File , 12 KB - mhealth_v10i4e35626_app2.docx]

Multimedia Appendix 3
Flow diagram.
[DOCX File , 43 KB - mhealth_v10i4e35626_app3.docx]

Multimedia Appendix 4
Supplementary tables.
[DOCX File , 24 KB - mhealth_v10i4e35626_app4.docx]

References

1. Hallal PC, Bauman AE, Heath GW, Kohl 3rd HW, Lee IM, Pratt M. Physical activity: more of the same is not enough. Lancet 2012;380(9838):190-191. [doi: 10.1016/S0140-6736(12)61027-7] [Medline: 22818932]
2. Lee I, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT, Lancet Physical Activity Series Working Group. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet 2012;380(9838):219-229 [FREE Full text] [doi: 10.1016/S0140-6736(12)61031-9] [Medline: 22818936]

3. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U, Lancet Physical Activity Series Working Group. Global physical activity levels: surveillance progress, pitfalls, and prospects. Lancet 2012;380(9838):247-257. [doi: 10.1016/S0140-6736(12)60646-1] [Medline: 22818937]
4. Wareham NJ, Rennie KL. The assessment of physical activity in individuals and populations: why try to be more precise about how physical activity is assessed? Int J Obes Relat Metab Disord 1998;22 Suppl 2:S30-S38. [Medline: 9778094]
5. Lynch J, Smith GD. A life course approach to chronic disease epidemiology. Annu Rev Public Health 2005;26:1-35. [doi: 10.1146/annurev.publhealth.26.021304.144505] [Medline: 15760279]
6. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. Br J Sports Med 2014;48(13):1019-1023 [FREE Full text] [doi: 10.1136/bjsports-2014-093546] [Medline: 24782483]
7. Van de Water AT, Holmes A, Hurley DA. Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography--a systematic review. J Sleep Res 2011;20(1 Pt 2):183-200 [FREE Full text] [doi: 10.1111/j.1365-2869.2009.00814.x] [Medline: 20374444]
8. Dobkin BH, Dorsch A. The promise of mHealth: daily activity monitoring and outcome assessments by wearable sensors. Neurorehabil Neural Repair 2011;25(9):788-798 [FREE Full text] [doi: 10.1177/1545968311425908] [Medline: 21989632]
9. Brage S, Westgate K, Franks PW, Stegle O, Wright A, Ekelund U, et al. Estimation of free-living energy expenditure by heart rate and movement sensing: a doubly-labelled water study. PLoS One 2015;10(9):e0137206 [FREE Full text] [doi: 10.1371/journal.pone.0137206] [Medline: 26349056]
10. Brage S, Ekelund U, Brage N, Hennings MA, Froberg K, Franks PW, et al. Hierarchy of individual calibration levels for heart rate and accelerometry to measure physical activity. J Appl Physiol (1985) 2007;103(2):682-692 [FREE Full text] [doi: 10.1152/japplphysiol.00092.2006] [Medline: 17463305]
11. Brage S, Brage N, Franks PW, Ekelund U, Wong MY, Andersen LB, et al. Branched equation modeling of simultaneous accelerometry and heart rate monitoring improves estimate of directly measured physical activity energy expenditure. J Appl Physiol (1985) 2004;96(1):343-351 [FREE Full text] [doi: 10.1152/japplphysiol.00703.2003] [Medline: 12972441]
12. Thompson D, Batterham AM, Bock S, Robson C, Stokes K. Assessment of low-to-moderate intensity physical activity thermogenesis in young adults using synchronized heart rate and accelerometry with branched-equation modeling. J Nutr 2006;136(4):1037-1042. [doi: 10.1093/jn/136.4.1037] [Medline: 16549471]
13. Mercer K, Li M, Giangregorio L, Burns C, Grindrod K. Behavior change techniques present in wearable activity trackers: a critical analysis. JMIR Mhealth Uhealth 2016;4(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4461] [Medline: 27122452]
14. Phillips SM, Cadmus-Bertram L, Rosenberg D, Buman MP, Lynch BM. Wearable technology and physical activity in chronic disease: opportunities and challenges. Am J Prev Med 2018;54(1):144-150 [FREE Full text] [doi: 10.1016/j.amepre.2017.08.015] [Medline: 29122356]
15. Dooley EE, Golaszewski NM, Bartholomew JB. Estimating accuracy at exercise intensities: a comparative study of self-monitoring heart rate and physical activity wearable devices. JMIR Mhealth Uhealth 2017;5(3):e34 [FREE Full text] [doi: 10.2196/mhealth.7043] [Medline: 28302596]
16. Gorny AW, Liew SJ, Tan CS, Müller-Riemenschneider F. Fitbit charge HR wireless heart rate monitor: validation study conducted under free-living conditions. JMIR Mhealth Uhealth 2017;5(10):e157 [FREE Full text] [doi: 10.2196/mhealth.8233] [Medline: 29055881]
17. O'Driscoll R, Turicchi J, Beaulieu K, Scott S, Matu J, Deighton K, et al. How well do activity monitors estimate energy expenditure? A systematic review and meta-analysis of the validity of current technologies. Br J Sports Med 2020;54(6):332-340. [doi: 10.1136/bjsports-2018-099643] [Medline: 30194221]
18. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015;12:159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
19. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR Mhealth Uhealth 2020;8(9):e18694 [FREE Full text] [doi: 10.2196/18694] [Medline: 32897239]
20. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Yoo JY, et al. Accuracy of Fitbit devices: systematic review and narrative syntheses of quantitative data. JMIR Mhealth Uhealth 2018;6(8):e10527 [FREE Full text] [doi: 10.2196/10527] [Medline: 30093371]
21. Hills AP, Mokhtar N, Byrne NM. Assessment of physical activity and energy expenditure: an overview of objective measures. Front Nutr 2014;1:5 [FREE Full text] [doi: 10.3389/fnut.2014.00005] [Medline: 25988109]
22. Horton JF, Stergiou P, Fung TS, Katz L. Comparison of Polar M600 optical heart rate and ECG heart rate during exercise. Med Sci Sports Exerc 2017;49(12):2600-2607. [doi: 10.1249/MSS.0000000000001388] [Medline: 29135785]
23. O'Connell S, ÓLaighin G, Quinlan LR. When a step is not a step! Specificity analysis of five physical activity monitors. PLoS One 2017;12(1):e0169616 [FREE Full text] [doi: 10.1371/journal.pone.0169616] [Medline: 28085918]
24. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015;4(1):1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]

25. Supplemental materials for preprint: accuracy and precision of energy expenditure, heart rate, and steps measured by combined-sensing Fitbits: protocol for a systematic review and meta-analyses. Open Science Framework. 2020. URL: https://osf.io/ky4jp/ [accessed 2022-03-21]
26. Preston L, Carroll C, Gardois P, Paisley S, Kaltenthaler E. Improving search efficiency for systematic reviews of diagnostic test accuracy: an exploratory study to assess the viability of limiting to MEDLINE, EMBASE and reference checking. Syst Rev 2015;4:82 [FREE Full text] [doi: 10.1186/s13643-015-0074-7] [Medline: 26113080]
27. Tipton E, Shuster J. A framework for the meta-analysis of Bland-Altman studies based on a limits of agreement approach. Stat Med 2017;36(23):3621-3635 [FREE Full text] [doi: 10.1002/sim.7352] [Medline: 28664537]
28. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7(3):177-188. [doi: 10.1016/0197-2456(86)90046-2] [Medline: 3802833]
29. Scammacca N, Roberts G, Stuebing KK. Meta-analysis with complex research designs: dealing with dependence from multiple measures and multiple group comparisons. Rev Educ Res 2014;84(3):328-364 [FREE Full text] [doi: 10.3102/0034654313500826] [Medline: 25309002]
30. Tanner-Smith EE, Tipton E. Robust variance estimation with dependent effect sizes: practical considerations including a software tutorial in Stata and spss. Res Synth Methods 2014;5(1):13-30. [doi: 10.1002/jrsm.1091] [Medline: 26054023]
31. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010;19(4):539-549 [FREE Full text] [doi: 10.1007/s11136-010-9606-8] [Medline: 20169472]
32. Lu MJ, Zhong WH, Liu YX, Miao HZ, Li YC, Ji MH. Sample size for assessing agreement between two methods of measurement by Bland-Altman method. Int J Biostat 2016;12(2):20150039. [doi: 10.1515/ijb-2015-0039] [Medline: 27838682]
33. Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiol Meas 2007;28(3):R1-39. [doi: 10.1088/0967-3334/28/3/R01] [Medline: 17322588]
34. Al-Kaisey AM, Koshy AN, Ha FJ, Spencer R, Toner L, Sajeev JK, et al. Accuracy of wrist-worn heart rate monitors for rate control assessment in atrial fibrillation. Int J Cardiol 2020;300:161-164. [doi: 10.1016/j.ijcard.2019.11.120] [Medline: 31787389]
35. Baek S, Ha Y, Park HW. Accuracy of wearable devices for measuring heart rate during conventional and Nordic walking. PM R 2021;13(4):379-386. [doi: 10.1002/pmrj.12424] [Medline: 32449291]
36. Bai Y, Hibbing P, Mantis C, Welk GJ. Comparative evaluation of heart rate-based monitors: Apple Watch vs Fitbit Charge HR. J Sports Sci 2018;36(15):1734-1741. [doi: 10.1080/02640414.2017.1412235] [Medline: 29210326]
37. Bai Y, Tompkins C, Gell N, Dione D, Zhang T, Byun W. Comprehensive comparison of Apple Watch and Fitbit monitors in a free-living setting. PLoS One 2021;16(5):e0251975 [FREE Full text] [doi: 10.1371/journal.pone.0251975] [Medline: 34038458]
38. Benedetto S, Caldato C, Bazzan E, Greenwood DC, Pensabene V, Actis P. Assessment of the Fitbit Charge 2 for monitoring heart rate. PLoS One 2018;13(2):e0192691 [FREE Full text] [doi: 10.1371/journal.pone.0192691] [Medline: 29489850]
39. Boudreaux BD, Hebert EP, Hollander DB, Williams BM, Cormier CL, Naquin MR, et al. Validity of wearable activity monitors during cycling and resistance exercise. Med Sci Sports Exerc 2018;50(3):624-633. [doi: 10.1249/MSS.0000000000001471] [Medline: 29189666]
40. Bunn JA, Jones C, Oliviera A, Webster MJ. Assessment of step accuracy using the consumer technology association standard. J Sports Sci 2019;37(3):244-248. [doi: 10.1080/02640414.2018.1491941] [Medline: 29958058]
41. Burton E, Hill KD, Lautenschlager NT, Thøgersen-Ntoumani C, Lewin G, Boyle E, et al. Reliability and validity of two fitness tracker devices in the laboratory and home environment for older community-dwelling people. BMC Geriatr 2018;18(1):103 [FREE Full text] [doi: 10.1186/s12877-018-0793-4] [Medline: 29724191]
42. Cadmus-Bertram L, Gangnon R, Wirkus EJ, Thraen-Borowski KM, Gorzelitz-Liebhauser J. The accuracy of heart rate monitoring by some wrist-worn activity trackers. Ann Intern Med 2017;166(8):610-612 [FREE Full text] [doi: 10.7326/L16-0353] [Medline: 28395305]
43. Chow JJ, Thom JM, Wewege MA, Ward RE, Parmenter BJ. Accuracy of step count measured by physical activity monitors: the effect of gait speed and anatomical placement site. Gait Posture 2017;57:199-203. [doi: 10.1016/j.gaitpost.2017.06.012] [Medline: 28666177]
44. Chowdhury A, Sehgal S, Rabih F, Gadre A, Highland KB. Accuracy of accelerometer (Fitbit® ChargeHR™) measured distance and heart rate in patients with pulmonary arterial hypertension. Am J Respir Crit Care Med 2017;195:A3116.
45. Claes J, Avila A, Cornelissen V, Cool H, Segers A, Buys R. Validity of three consumer-based physical activity devices: Fitbit charge HR, Garmin forerunner 225 and Mio fuse. Eur Heart J 2016;37:334-335.
46. Herkert CC, Kraal JJ, Van Loon EM, Van Hooff M, Brouwers RW, Kemps HM. Validation of two wrist-worn devices for the assessment of energy expenditure in patients with chronic heart failure and coronary artery disease. Eur J Prev Cardiol 2019;26(1_suppl):S7.
47. Düking P, Giessing L, Frenkel MO, Koehler K, Holmberg HC, Sperlich B. Wrist-worn wearables for monitoring heart rate and energy expenditure while sitting or performing light-to-vigorous physical activity: validation study. JMIR Mhealth Uhealth 2020;8(5):e16716 [FREE Full text] [doi: 10.2196/16716] [Medline: 32374274]

48. Etiwy M, Akhrass Z, Gillinov L, Alashi A, Wang R, Blackburn G, et al. Accuracy of wearable heart rate monitors in cardiac rehabilitation. Cardiovasc Diagn Ther 2019;9(3):262-271 [FREE Full text] [doi: 10.21037/cdt.2019.04.08] [Medline: 31275816]
49. Falgoust BJ, Handwerger B, Rouly L, Blaha O, Pontiff M, Eason JM. Accuracy of wearable activity monitors in the measurement of step count and distance. Cardiopulm Phys Ther J 2018;29(1):35-36.
50. Fokkema T, Kooiman TJ, Krijnen WP, Van Der Schans CP, DE Groot M. Reliability and validity of ten consumer activity trackers depend on walking speed. Med Sci Sports Exerc 2017;49(4):793-800. [doi: 10.1249/MSS.0000000000001146] [Medline: 28319983]
51. Gaynor M, Sawyer A, Jenkins S, Wood J. Variable agreement between wearable heart rate monitors during exercise in cystic fibrosis. ERJ Open Res 2019;5(4):00006-2019 [FREE Full text] [doi: 10.1183/23120541.00006-2019] [Medline: 31687372]
52. Gillinov S, Etiwy M, Wang R, Blackburn G, Phelan D, Gillinov AM, et al. Variable accuracy of wearable heart rate monitors during aerobic exercise. Med Sci Sports Exerc 2017;49(8):1697-1703. [doi: 10.1249/MSS.0000000000001284] [Medline: 28709155]
53. Jagim AR, Koch-Gallup N, Camic CL, Kroening L, Nolte C, Schroeder C, et al. The accuracy of fitness watches for the measurement of heart rate and energy expenditure during moderate intensity exercise. J Sports Med Phys Fitness 2021;61(2):205-211. [doi: 10.23736/S0022-4707.20.11151-4] [Medline: 32734757]
54. Jo E, Lewis K, Directo D, Kim MJ, Dolezal BA. Validation of biofeedback wearables for photoplethysmographic heart rate tracking. J Sports Sci Med 2016;15(3):540-547 [FREE Full text] [Medline: 27803634]
55. Lai B, Sasaki JE, Jeng B, Cederberg KL, Bamman MM, Motl RW. Accuracy and precision of three consumer-grade motion sensors during overground and treadmill walking in people with Parkinson disease: cross-sectional comparative study. JMIR Rehabil Assist Technol 2020;7(1):e14059 [FREE Full text] [doi: 10.2196/14059] [Medline: 31944182]
56. Lamont RM, Daniel HL, Payne CL, Brauer SG. Accuracy of wearable physical activity trackers in people with Parkinson's disease. Gait Posture 2018;63:104-108. [doi: 10.1016/j.gaitpost.2018.04.034] [Medline: 29729611]
57. Lee JM, An HS, Kang SK, Kim Y. Examining the validity of Fitbit charge HR TM for measuring heart rate in free-living conditions. Res Presentations 2016:15.
58. Modave F, Guo Y, Bian J, Gurka MJ, Parish A, Smith MD, et al. Mobile device accuracy for step counting across age groups. JMIR Mhealth Uhealth 2017;5(6):e88 [FREE Full text] [doi: 10.2196/mhealth.7870] [Medline: 28659255]
59. Montes J, Tandy R, Young J, Lee SP, Navalta JW. Step count reliability and validity of five wearable technology devices while walking and jogging in both a free motion setting and on a treadmill. Int J Exerc Sci 2020;13(7):410-426 [FREE Full text] [Medline: 32509137]
60. Montoye AH, Mitrzyk JR, Molesky MJ. Comparative accuracy of a wrist-worn activity tracker and a smart shirt for physical activity assessment. Meas Phys Educ Exerc Sci 2017;21(4):201-211. [doi: 10.1080/1091367X.2017.1331166]
61. Morris CE, Wessel PA, Tinius RA, Schafer MA, Maples JM. Validity of activity trackers in estimating energy expenditure during high-intensity functional training. Res Q Exerc Sport 2019;90(3):377-384. [doi: 10.1080/02701367.2019.1603989] [Medline: 31046630]
62. Muggeridge DJ, Hickson K, Davies AV, Giggins OM, Megson IL, Gorely T, et al. Measurement of heart rate using the polar OH1 and Fitbit charge 3 wearable devices in healthy adults during light, moderate, vigorous, and sprint-based exercise: validation study. JMIR Mhealth Uhealth 2021;9(3):e25313 [FREE Full text] [doi: 10.2196/25313] [Medline: 33764310]
63. Nelson BW, Allen NB. Accuracy of consumer wearable heart rate measurement during an ecologically valid 24-hour period: intraindividual validation study. JMIR Mhealth Uhealth 2019;7(3):e10828 [FREE Full text] [doi: 10.2196/10828] [Medline: 30855232]
64. Nuss KJ, Thomson EA, Courtney JB, Comstock A, Reinwald S, Blake S, et al. Assessment of accuracy of overall energy expenditure measurements for the Fitbit Charge HR 2 and Apple Watch. Am J Health Behav 2019;43(3):498-505. [doi: 10.5993/AJHB.43.3.5] [Medline: 31046881]
65. Pasadyn SR, Soudan M, Gillinov M, Houghtaling P, Phelan D, Gillinov N, et al. Accuracy of commercially available heart rate monitors in athletes: a prospective study. Cardiovasc Diagn Ther 2019;9(4):379-385 [FREE Full text] [doi: 10.21037/cdt.2019.06.05] [Medline: 31555543]
66. Powierza CS, Clark MD, Hughes JM, Carneiro KA, Mihalik JP. Validation of a self-monitoring tool for use in exercise therapy. PM R 2017;9(11):1077-1084 [FREE Full text] [doi: 10.1016/j.pmrj.2017.03.012] [Medline: 28400221]
67. Pribyslavska V, Caputo JL, Coons JM, Barry VW. Impact of EPOC adjustment on estimation of energy expenditure using activity monitors. J Med Eng Technol 2018;42(4):265-273. [doi: 10.1080/03091902.2018.1472823] [Medline: 29911930]
68. Reddy RK, Pooni R, Zaharieva DP, Senf B, El Youssef J, Dassau E, et al. Accuracy of wrist-worn activity monitors during common daily physical activities and types of structured exercise: evaluation study. JMIR Mhealth Uhealth 2018;6(12):e10338 [FREE Full text] [doi: 10.2196/10338] [Medline: 30530451]
69. Salazar VE, Lucio ND, Funk MD. Accuracy of Fitbit charge 2 worn at different wrist locations during exercise. Int J Exerc Sci 2017;2(9):41.

70. Shcherbina A, Mattsson CM, Waggott D, Salisbury H, Christle JW, Hastie T, et al. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. J Pers Med 2017;7(2):3 [FREE Full text] [doi: 10.3390/jpm7020003] [Medline: 28538708]
71. Siddall AG, Powell SD, Needham-Beck SC, Edwards VC, Thompson JE, Kefyalew SS, et al. Validity of energy expenditure estimation methods during 10 days of military training. Scand J Med Sci Sports 2019;29(9):1313-1321. [doi: 10.1111/sms.13488] [Medline: 31136027]
72. Sjöberg V, Westergren J, Monnier A, Lo Martire R, Hagströmer M, Äng BO, et al. Wrist-worn activity trackers in laboratory and free-living settings for patients with chronic pain: criterion validity study. JMIR Mhealth Uhealth 2021;9(1):e24806 [FREE Full text] [doi: 10.2196/24806] [Medline: 33433391]
73. Stahl SE, An HS, Dinkel DM, Noble JM, Lee JM. How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? BMJ Open Sport Exerc Med 2016;2(1):e000106 [FREE Full text] [doi: 10.1136/bmjsem-2015-000106] [Medline: 27900173]
74. Tam KM, Cheung SY. Validation of electronic activity monitor devices during treadmill walking. Telemed J E Health 2018;24(10):782-789. [doi: 10.1089/tmj.2017.0263] [Medline: 29364065]
75. Tedesco S, Sica M, Ancillao A, Timmons S, Barton J, O'Flynn B. Validity evaluation of the Fitbit Charge2 and the Garmin vivosmart HR+ in free-living environments in an older adult cohort. JMIR Mhealth Uhealth 2019;7(6):e13084 [FREE Full text] [doi: 10.2196/13084] [Medline: 31219048]
76. Thiebaud RS, Funk MD, Patton JC, Massey BL, Shay TE, Schmidt MG, et al. Validity of wrist-worn consumer products to measure heart rate and energy expenditure. Digit Health 2018;4:2055207618770322 [FREE Full text] [doi: 10.1177/2055207618770322] [Medline: 29942628]
77. Thomson EA, Nuss K, Comstock A, Reinwald S, Blake S, Pimentel RE, et al. Heart rate measures from the Apple Watch, Fitbit Charge HR 2, and electrocardiogram across different exercise intensities. J Sports Sci 2019;37(12):1411-1419. [doi: 10.1080/02640414.2018.1560644] [Medline: 30657025]
78. Tophøj KH, Petersen MG, Sæbye C, Baad-Hansen T, Wagner S. Validity and reliability evaluation of four commercial activity trackers' step counting performance. Telemed J E Health 2018;24(9):669-677. [doi: 10.1089/tmj.2017.0264] [Medline: 29303680]
79. Wahl Y, Düking P, Droszez A, Wahl P, Mester J. Criterion-validity of commercially available physical activity tracker to estimate step count, covered distance and energy expenditure during sports conditions. Front Physiol 2017;8:725 [FREE Full text] [doi: 10.3389/fphys.2017.00725] [Medline: 29018355]
80. Wallen MP, Gomersall SR, Keating SE, Wisløff U, Coombes JS. Accuracy of heart rate watches: implications for weight management. PLoS One 2016;11(5):e0154420 [FREE Full text] [doi: 10.1371/journal.pone.0154420] [Medline: 27232714]
81. Wang R, Blackburn G, Desai M, Phelan D, Gillinov L, Houghtaling P, et al. Accuracy of wrist-worn heart rate monitors. JAMA Cardiol 2017;2(1):104-106. [doi: 10.1001/jamacardio.2016.3340] [Medline: 27732703]
82. Xie J, Wen D, Liang L, Jia Y, Gao L, Lei J. Evaluating the validity of current mainstream wearable devices in fitness tracking under various physical activities: comparative study. JMIR Mhealth Uhealth 2018;6(4):e94 [FREE Full text] [doi: 10.2196/mhealth.9754] [Medline: 29650506]
83. Zhang P, Burns RD, Fu Y, Godin S, Byun W. Agreement between the Apple Series 1, LifeTrak Core C200, and Fitbit Charge HR with indirect calorimetry for assessing treadmill energy expenditure. Int J Environ Res Public Health 2019;16(20):3812 [FREE Full text] [doi: 10.3390/ijerph16203812] [Medline: 31658628]
84. American National Standard. Cardiac monitors, heart rate meters, and alarms. Association for the Advancement of Medical Instrumentation. 2002. URL: https://docplayer.net/34982183-Aami-american-national-standard-cardiac-monitors-heart-rate-meters-and-alarms-ansi-aami-ec13-2002.html [accessed 2021-08-17]
85. Slaght J, Sénéchal M, Hrubeniuk TJ, Mayo A, Bouchard DR. Walking cadence to exercise at moderate intensity for adults: a systematic review. J Sports Med (Hindawi Publ Corp) 2017;2017:4641203 [FREE Full text] [doi: 10.1155/2017/4641203] [Medline: 28459099]
86. Physical activity for a healthy weight. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/healthyweight/physical_activity/index.html [accessed 2021-08-17]
87. Welk GJ, Bai Y, Lee JM, Godino J, Saint-Maurice PF, Carr L. Standardizing analytic methods and reporting in activity monitor validation studies. Med Sci Sports Exerc 2019;51(8):1767-1780 [FREE Full text] [doi: 10.1249/MSS.0000000000001966] [Medline: 30913159]

Abbreviations

bpm: beats per minute
COSMIN: Consensus-Based Standards for the Selection of Health Measurement Instruments
PRISMA-P: Preferred Reporting Items for Systematic Review and Meta-Analyses Protocols
PROSPERO: International Prospective Register of Systematic Reviews
RVE: robust variance estimation

Edited by L Buis; submitted 14.12.21; peer-reviewed by J Sanders; comments to author 12.01.22; revised version received 27.01.22; accepted 10.02.22; published 13.04.22.

Please cite as:
Chevance G, Golaszewski NM, Tipton E, Hekler EB, Buman M, Welk GJ, Patrick K, Godino JG
Accuracy and Precision of Energy Expenditure, Heart Rate, and Steps Measured by Combined-Sensing Fitbits Against Reference Measures: Systematic Review and Meta-analysis
JMIR Mhealth Uhealth 2022;10(4):e35626
URL: https://mhealth.jmir.org/2022/4/e35626
doi:10.2196/35626
PMID:35416777

Review

Digital Health Technologies for Long-term Self-management of Osteoporosis: Systematic Review and Meta-analysis

Ghada Alhussein[1,2*], BSc, MSc; Leontios Hadjileontiadis[1,2,3*], BSc, MSc, PhD

[1]Department of Biomedical Engineering, Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates

[2]Healthcare Innovation Center, Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates

[3]Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece

[*]all authors contributed equally

Corresponding Author:
Ghada Alhussein, BSc, MSc
Department of Biomedical Engineering
Khalifa University of Science and Technology
AlSaada Street
Abu Dhabi
United Arab Emirates
Phone: 971 508292048
Email: ghada.alhussein@ku.ac.ae

Abstract

Background: Osteoporosis is the fourth most common chronic disease worldwide. The adoption of preventative measures and effective self-management interventions can help improve bone health. Mobile health (mHealth) technologies can play a key role in the care and self-management of patients with osteoporosis.

Objective: This study presents a systematic review and meta-analysis of the currently available mHealth apps targeting osteoporosis self-management, aiming to determine the current status, gaps, and challenges that future research could address, as well as propose appropriate recommendations.

Methods: A systematic review of all English articles was conducted, in addition to a survey of all apps available in iOS and Android app stores as of May 2021. A comprehensive literature search (2010 to May 2021) of PubMed, Scopus, EBSCO, Web of Science, and IEEE Xplore was conducted. Articles were included if they described apps dedicated to or useful for osteoporosis (targeting self-management, nutrition, physical activity, and risk assessment) delivered on smartphone devices for adults aged ≥18 years. Of the 32 articles, a random effects meta-analysis was performed on 13 (41%) studies of randomized controlled trials, whereas the 19 (59%) remaining studies were only included in the narrative synthesis as they did not provide enough data.

Results: In total, 3906 unique articles were identified. Of these 3906 articles, 32 (0.81%) articles met the inclusion criteria and were reviewed in depth. The 32 studies comprised 14,235 participants, of whom, on average, 69.5% (n=9893) were female, with a mean age of 49.8 (SD 17.8) years. The app search identified 23 relevant apps for osteoporosis self-management. The meta-analysis revealed that mHealth-supported interventions resulted in a significant reduction in pain (Hedges g −1.09, 95% CI −1.68 to −0.45) and disability (Hedges g −0.77, 95% CI −1.59 to 0.05). The posttreatment effect of the digital intervention was significant for physical function (Hedges g 2.54, 95% CI −4.08 to 4.08) but nonsignificant for well-being (Hedges g 0.17, 95% CI −1.84 to 2.17), physical activity (Hedges g 0.09, 95% CI −0.59 to 0.50), anxiety (Hedges g −0.29, 95% CI −6.11 to 5.53), fatigue (Hedges g −0.34, 95% CI −5.84 to 5.16), calcium (Hedges g −0.05, 95% CI −0.59 to 0.50), vitamin D intake (Hedges g 0.10, 95% CI −4.05 to 4.26), and trabecular score (Hedges g 0.06, 95% CI −1.00 to 1.12).

Conclusions: Osteoporosis apps have the potential to support and improve the management of the disease and its symptoms; they also appear to be valuable tools for patients and health professionals. However, most of the apps that are currently available lack clinically validated evidence of their efficacy and focus on a limited number of symptoms. A more holistic and personalized approach within a cocreation design ecosystem is needed.

Trial Registration: PROSPERO 2021 CRD42021269399; https://tinyurl.com/2sw454a9

(JMIR Mhealth Uhealth 2022;10(4):e32557) doi:10.2196/32557

KEYWORDS

mHealth; digital health; osteoporosis; self-management; systematic review; meta-analysis; chronic disease; bone health; self-management; nutrition; physical activity; risk assessment; mobile phone

Introduction

Background

Osteoporosis, or porous bone, is a serious chronic disease in which the density of bones is silently and progressively reduced, resulting in a more porous and fragile structure [1]. This disease takes a huge personal and economic toll on the world [2]. The disabilities caused by osteoporosis outweigh those caused by cancer and many other chronic diseases. Both men and women can develop osteoporosis; however, women are more susceptible to this disease [3]. This silent killer is estimated to affect 200 million women worldwide—approximately one-tenth of women aged 60 years, one-fifth of women aged 70 years, two-fifths of women aged 80 years, and two-thirds of women aged 90 years [4]. By 2050, the worldwide incidence of hip fractures in both men and women is projected to increase significantly compared with the current number of cases [5]. Moreover, it is suggested that most individuals at high risk of osteoporosis are not properly diagnosed and are neither identified nor treated. It is also noted that >40% of patients with osteoporosis drop out from exercise therapies [6], and between 40% and 70% of patients adhere to drug therapies [7], which is not the case in patients with cancer or cardiovascular diseases. Therefore, it is important to identify and treat patients at risk of fracture, not only by prescribing effective medications but also by equipping them with the information they need to take appropriate behavior to prevent the consequences of the disease. This will substantially reduce the long-term burden of osteoporosis. Reducing the risk of the first fracture from 8% to 2% can reduce the 5-year fracture incidence from approximately 34% to 10% [2,5].

The current landscape of a rapidly aging population, accompanied by multiple chronic conditions, presents numerous challenges to optimally supporting the complex needs of this group. Therefore, it is essential to find better and affordable alternatives to hospital and institutional care that can support older adults in their homes rather than moving them to health care providers. The use of health-related mobile apps, or mobile health (mHealth), has emerged as an important and useful tool for improving health outcomes in chronic disease self-management [8]. Self-management is a very effective factor that can enhance overall health; it encompasses tasks, such as goal setting, active motivation, self-monitoring, decision-making, problem solving, planning for and engaging in specific behaviors, self-evaluation, stress management and emotional regulation, coping with lapses and setbacks, and assertive communication [9]. These mHealth apps allow for effective communication between patients and physicians, better clinical decision-making, and improved patient outcomes. Moreover, mHealth apps can support people to manage their own health, promote healthy living, and have access to the necessary information when and where they need it. They also have a groundbreaking impact on the pharmaceutical and health care industry because of their faster, better, and cheaper health management benefits [10,11].

The number of apps available on the planet exceeds 8 million, of which 60% are available on both Android and iOS app stores [12]. As of 2017, there are 325,000 mHealth apps with an annual download of >3.7 billion [13]. This increase in demand has resulted mainly from the growing penetration of smartphones and the emergence of advanced technologies in the health care sector. Moreover, the adoption of mHealth is likely to increase further, especially because of COVID-19 [14] and in remote areas that lack hospitals and clinics [15,16]. Despite the increasing number of mHealth apps, a limited number have been dedicated to patients with osteoporosis, although it is a major worldwide health challenge. In addition, a few studies have focused on long-term self-management of osteoporosis, which extends throughout the patient's life. Even with the availability of cost-effective and well-tolerated treatments for osteoporosis, there is still no appropriate self-management of the disease to prevent fractures [17]. The individual responsibility for health and self-management of chronic diseases has been a concept with growing interest during the past decades [18], and mHealth can be useful for this purpose.

Objective

The motivation behind this systematic review stems from the fact that, to the best of our knowledge, there is no other review so far that explores mHealth apps dedicated to osteoporosis self-management available in both the web-based app market and in the research field. The present systematic review and meta-analysis were undertaken to come up with the identification of the current status of osteoporosis-related mHealth solutions, reveal any lack of functionalities, identify challenges and barriers, and propose recommendations for more personalized and effective remote health care monitoring and interventions. In this way, efforts toward the development and testing of a holistic mobile app to support patients at risk of or with osteoporosis are better informed. Osteoporosis self-management apps with a holistic approach should comprise a wide variety of features, including nutrition, physical exercise, medication, and performance monitoring, in addition to involving a wide spectrum of stakeholders, from rheumatologists to other health care professionals, and requiring patients to be well-informed and to take an active role in their own car, while providing an incentive for physicians to trust, integrate, and implement mHealth apps into their medical practice.

Methods

Data Sources and Searches

For this systematic review, published sources were identified by searching PubMed, Scopus, Web of Science, IEEE Xplore, and EBSCO databases. A comprehensive combination of keywords was used to have the maximum possible coverage: *Osteoporosis* AND *Technology* OR *mHealth* OR *eHealth* OR *Remote Care* OR *Digital health technologies* OR *smartphone* OR *mobile phone* OR *Mobile applications* OR *app* or *Self-Management* OR *Disease management* OR *Bone health*.

The titles and abstracts of all records were examined, whereas the full text was screened only for the potentially relevant studies for final inclusion; any duplicates were removed. Table S1 in Multimedia Appendix 1 shows the search terms and results yielded from the different databases in detail.

Study Selections

The inclusion criteria were original studies or research papers, including people (both male and female) aged ≥18 years with no mental health conditions. The selected studies evaluated digital health technology, primarily designed to support targeted patient communication, education, diagnosis, real-time monitoring, and empowerment in the form of mobile phone apps supported by other audiovisual technologies. Moreover, we considered studies that use intelligent wireless sensors to capture any critical vital signs to support patients with osteoporosis in the long-term self-management of the disease. We also included studies that proposed a design or framework for mHealth apps targeting patients with osteoporosis. As no mHealth apps dedicated to osteoporosis self-management were found before 2010, only full-text studies published in peer-reviewed journals and in English from January 2010 to May 2021 were included.

Studies with participants who had mental disorders were excluded. In addition to studies that did not have full text available, we eliminated reviews, posters, letters, and expert opinion publications. We also did not consider studies with technological interventions not targeting or not useful for osteoporosis self-management, those that had no clear relationship with osteoporosis, those related to other musculoskeletal conditions, or those not useful for osteoporosis. Articles that did not use mobile apps were excluded, in addition to studies that examined social network platforms and services (such as Telegram, Skype, WhatsApp, or Facebook), emails, and the web. In the same context, we excluded studies that did not use any mobile app or use mobile technologies as an auxiliary tool, namely, by sending SMS text messages to engage patients in certain activities or behaviors.

mHealth App Selection

We searched for *osteoporosis*, *bone health*, and *fracture* in different web-based app stores, including Google Play Store, Apple Store (iTunes), Amazon App store, Samsung Galaxy store, and GetJar. We found 72 apps, among which we selected only apps that were in English, targeted patients with osteoporosis, and focused on health, fitness, nutrition, and health categories. We excluded apps that were in the games and entertainment categories, apps that only recorded users' data without any feedback, and apps that provided access to magazine conferences or journals. The remaining apps were categorized according to the main features they provide, such as educational content, prediction and assessment tools, and users' tracking of osteoporosis-related pain and symptoms or both.

Data Extraction and Quality Assessment

The following information was abstracted from each study: sample size, sample age range, app name, app purpose, app operating platform, study design, intervention period, and major outcome indices. Publication bias of randomized controlled trial (RCT) studies was evaluated using the Cochrane risk of bias (ROB; version 2.0) tool [19], whereas the bias of the nonrandomized comparative studies was assessed using the ROB In Nonrandomized Studies of Interventions tool [20]. The latter comprises 7 domains to assess bias because of confounding factors, selection of participants, classification of interventions, deviation from intended interventions, missing data, measurement of outcomes, and selection of reported results. The ROB adjudications are categorized with their corresponding color schemes as follows: low risk (green), moderate risk (yellow), serious risk (orange), critical risk (red), or no information (gray).

The selection, screening, data abstraction, and quality appraisals were performed by 2 reviewers (GA and LH). Any disagreements between the reviewers were resolved through discussion.

Data Synthesis and Statistical Analysis (Meta-analysis)

Data from 41% (13/32) of studies were pooled in a statistical meta-analysis using meta-essential [21]. A random effects model was performed for 10 outcomes to compare before and after mHealth app use; that is, calcium intake (mg per day), vitamin D intake (µg per day), bone mineral density (BMD) levels, physical activity (hours per week), pain intensity, disability, physical functioning, well-being, fatigue, and anxiety. For each included outcome, the Hedges adjusted g [22] effect size was calculated and reported with 95% CIs. Heterogeneity was statistically assessed using Cochrane Q [23] and I^2 tests, with high values (I^2>50%) indicating high heterogeneity [24]. A 2-tailed P<.05 was considered significant in all the analyses. Statistical analyses were performed using random effects models. Moreover, statistical findings from the remaining 59% (19/32) of studies included in the review were narratively interpreted.

On the basis of the features provided by the apps, a scoring system was created for each app from the web-based market and those included from the research field. The selection of the scoring features stemmed from a combination of related theories. In particular, we followed the Technology Acceptance Model [25], which emphasizes the key factors that predict technology adoption by an individual based on the perceived usefulness and ease of use of the related technology, in our case, mHealth apps. Consequently, features that reflect the usefulness and ease of the proposed osteoporosis-related apps were considered. Moreover, features such as aesthetics and minimalistic design, recognition rather than recall, and error prevention were selected based on the 10 usability heuristics for user interface design of Nielsen [26,27], which were considered to be among the most frequent defects in mobile apps. Furthermore, many mHealth apps are used for consumption of a healthy diet, disease diagnosis, tracking physical activities, calculating calories, and monitoring sleep quality [28,29]. As these categories have already been accepted as interesting features in mHealth apps and are considered useful, we included them in the proposed analysis. Other features, such as notifications and reminders, create a sense of emotional bonding between users and the mHealth app, allowing them to keep self-managing their disease and, therefore, apply better monitoring of its symptoms and

current health status [28]. Data sharing with designated individuals is another important feature that users consider important in any mHealth app [30]. In this way, a holistic perspective of the necessary features that could benefit the usefulness, easiness, user engagement, and plan adherence was followed. Apparently, as the selected features were spread across the different apps and were evaluated by different end users, no weighting process was applied. This allowed for an objective basis of scoring analysis across all studies toward the maximization of the integration and competence of the apps' features. A feature weighting process would be useful if the focus of the analysis was placed on specific app functionalities or if an app could accommodate all features in an integrated way and be evaluated by end users, providing a rating of each feature's significance.

Table 1 provides an overview of the selected features and how they relate to self-management. For the web-based apps, 13 features were considered. These features were (1) diagnosis, (2) diet, (3) medication, (4) fractures, (5) error prevention (helping users recognize, diagnose, and recover from errors), (6) exercise, (7) visual aids, (8) data sharing, (9) social network, (10) reminders, (11) health warnings, (12) aesthetic and minimalistic design (avoid providing irrelevant or rarely required information), and (13) recognition rather than recall (remember user's choices and visible and easily retrievable instructions of use). Similarly, 13 features were evaluated for the research apps. These were (1) diet, (2) exercise, (3) diagnosis, (4) medication, (5) data sharing or export, (6) planning, (7) notifications, (8) chatbot, (9) visual aids, (10) progress tracking, (11) feedback, (12) communication, and (13) artificial intelligence (AI). The presence of each feature added 1 point to the total accumulated score for each app. The accumulated score (out of 13) was converted to a score out of 5. Finally, these scores were ranked in decreasing order (holistic osteoporosis management apps to atomistic apps). The raw cutoff for the selection of the selected studies or app was determined based on the mean score versus studies cumulative plot (Figures 1 and 2).

Table 1. Self-management features for both research and web-based apps.

Self-management facet	Web-based market app feature	Research app features
Socialization	• Networking capabilities • Data sharing	• Data sharing or export • Communication
Scheduling	• Reminders • Medication plan • Diet programs • Exercises	• Planning • Medication plan • Diet programs • Exercises
Warnings	• Fractures • Health warnings	• Notifications
User acceptability and usability	• Visual aids • Aesthetic and minimalistic designa • Recognition rather than recalla • Error prevention[a]	• Visual aids
Personalization or adaptation to change	N/A[b]	• Chatbot • Artificial intelligence
Performance monitoring	N/A	• Feedback • Progress tracking
Self-care	• Diagnosis	• Diagnosis

[a]These features were selected based on the 10 usability heuristics for user interface design of Nielsen and Mack [26], which were considered to be among the most frequent defects in mobile apps. It was not possible to evaluate some usability features in the research apps as they were not publicly available in app stores.

[b]N/A: not applicable.

Figure 1. Scoring for research apps: (A) Mean score per app available in the literature, with a raw cutoff score of 2.7; apps above the threshold provide a more holistic self-management plan. (B) Selected features with their mean score representing how often they were present in the apps. Features with the highest scores were available in a larger number of apps; features with the lowest scores (ie, chatbot and artificial intelligence) were present in only 1 app [31-59].

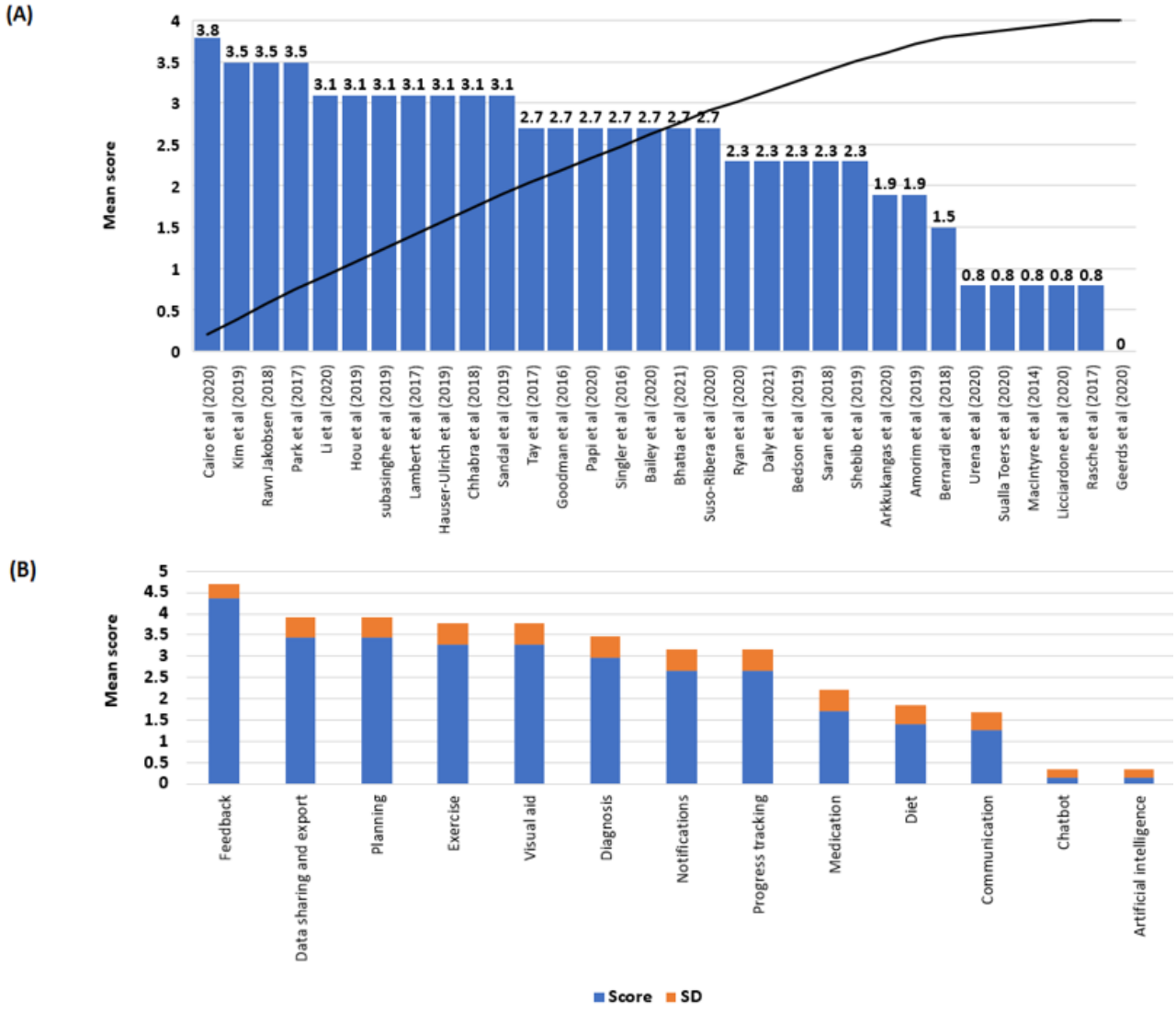
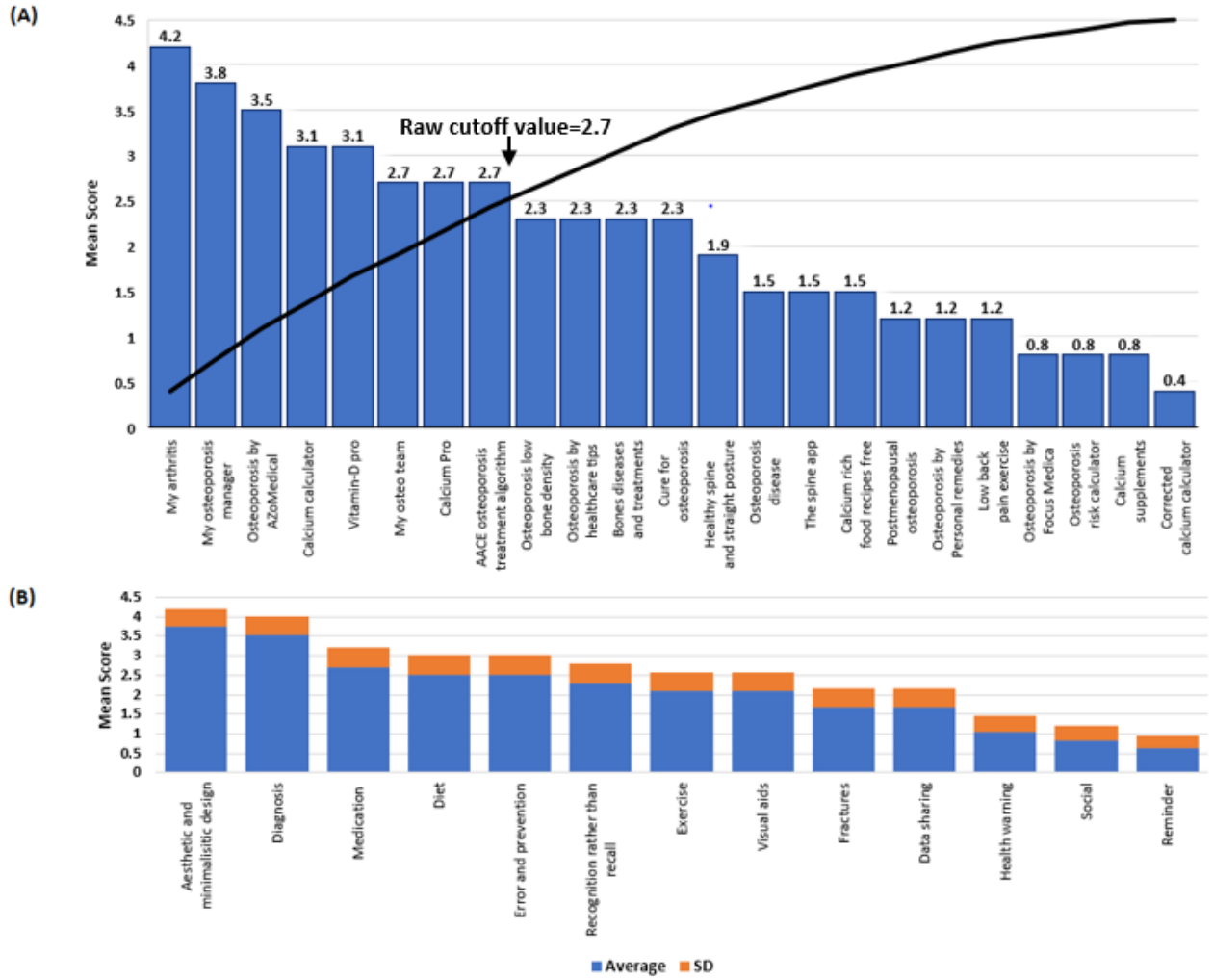


Figure 2. Web-based apps scores: (A) Mean score per app available in the web-based markets, with a raw cutoff score of 2.7; apps above the threshold provide a more holistic self-management plan. (B) Selected features with their mean score representing how often they were present in the apps. Features with the highest scores were available in more apps, whereas the features with the lowest scores were present in only 2 to 3 apps.



This systematic review was performed based on the recommendations of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [60]. The PRISMA checklist is provided in Multimedia Appendix 1, Table S2. The methods of analysis and the inclusion criteria were specified in advance.

Results

Literature Search Results

The literature search yielded 4185 articles, of which 3906 (93.33%) were screened. After removing duplicates and excluding studies on the basis of their titles and abstracts, 3.02% (118/3906) full texts were assessed for eligibility. In the final stage, 74.6% (88/118) of full-text citations did not meet the inclusion criteria. After completely reviewing the corresponding full-text articles, of the 88 articles, the total number of accepted articles was reduced to 32 (36%), of which 13 (41%) were selected for the meta-analysis. A PRISMA flowchart [60] for the article selection and exclusion process is provided in Figure 3A. We conducted an in-depth review of each of the included articles to classify them according to their research findings and determine their current state of knowledge. The data extracted from the selected papers are shown in Table 2.

Figure 3. Flow diagrams for the selection of (A) studies and (B) apps.

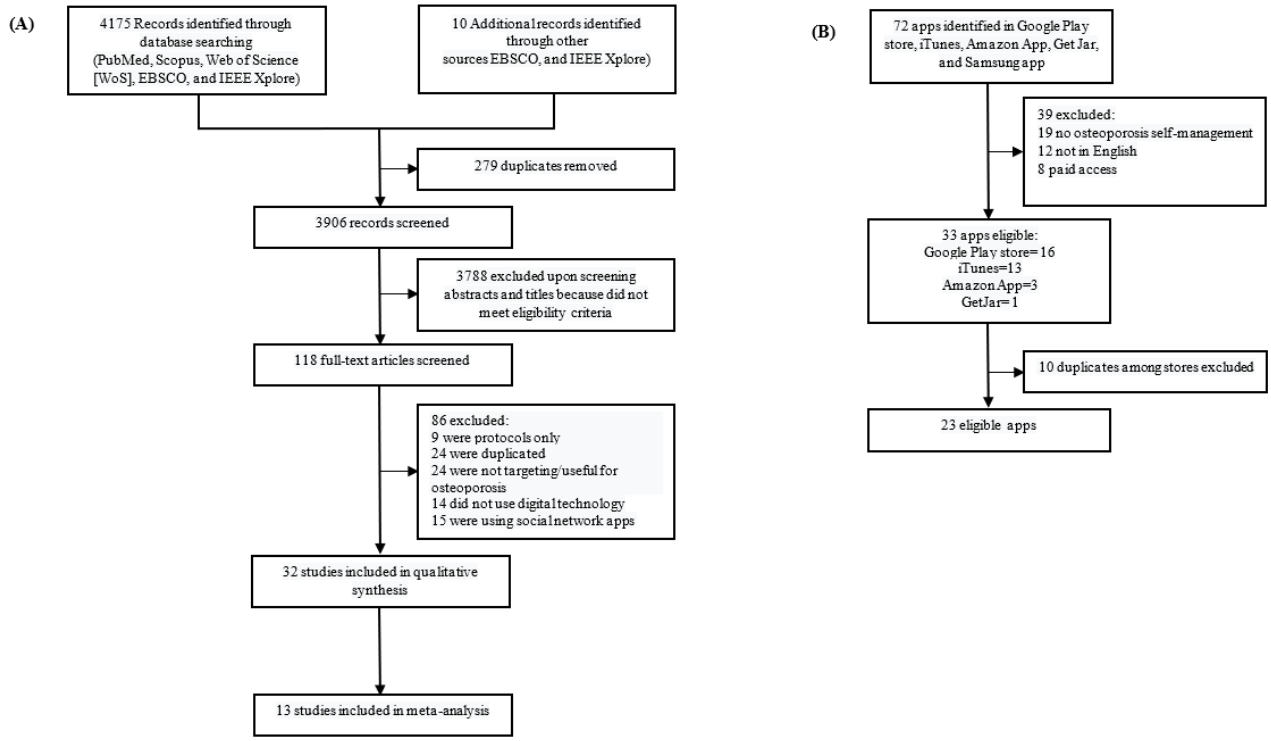


Table 2. Research app characteristics.

Author	App name	Sample size (age)	Experiment (participant sample size)	Platform[a] (private or public)	App purpose (direct or indirect)	Intervention period	Major outcome indices
Daly et al [38]	PhysiApp-patient portal	20 (>65 years)	App (20)	Android (public)	Remotely delivers and monitors an individually tailored, home-based multicomponent exercise program (indirect[b])	8 weeks	Feasibility, usability, physical activity enjoyment, changes in lower extremity function, and level of physical activity
Bhatia et al [44]	Manage My Pain	246 (mean age 57, SD 15 years)	App (111); no app (135)	Android and iOS (public)	Measures and monitors pain, function, and medication use (indirect)	92-183 days	Anxiety, depression, pain catastrophizing, satisfaction, daily opioid consumption, engagement
Cairo et al [59]	Vida app	127 (>18 years)	App (66); no app (61)	Android and iOS (public)	Improves wellness outcomes for survivors of breast cancer (indirect)	6 months	Physical activity, diary patterns, fatigue, and depression improvement
Hauser-Ulrich et al [53]	SELMA-Chatbot	102 (mean age 43.7 years)	App (59); no app (43)	Android and iOS	Promotes self-management of chronic pain (indirect)	12 weeks	Pain-related impairment, intention to change behavior, and pain intensity
Suso-Ribera et al [51]	Pain Monitor	87	App (43); no app (44)	N/A[c] (private)	Improves existent medical treatments for patients with chronic musculoskeletal pain (indirect)	4 weeks	Pain severity and interference, fatigue, depressed mood, anxiety, and anger
Licciardone et al [36]	N/A	102 (mean age 51 years)	App (52); no app (50)	N/A	Self-management of health - related quality of life (indirect)	3 months	Change in the SPADE[d] cluster score, changes in low back pain intensity, and back - related disability
Geerds et al [35]	N/A	24 (older adults >60 years)	App (24); no app (24)	N/A (private)	Monitors postoperative functional outcome after hip fracture (indirect)	12 and 18 weeks after surgery	Usability
Bailey et al [47]	Hinge Health app	10,264 (mean age 43.6 years)	App (10,264)	N/A (private)	Provides education, sensor-guided exercise therapy, and behavioral health support with one-on-one remote health coaching (indirect)	12 weeks	Pain measured by the Visual Analog Scale, engagement levels, program completion, program satisfaction, condition-specific pain measures, depression, anxiety, and work productivity
Ryan et al [48]	Striving app, Boning up	290 (40-60 years)	App (84); e-book (84); no app (84)	Android and iOS (private)	Provides information and feedback and monitors behavior change (direct[e])	12 months	Bone mineral density and trabecular bone scores
Papi et al [32]	Nymbl	35 (≥55 years)	App (35)	N/A (private)	Trains balance in the older population (indirect)	3 weeks for all, with optional follow-up for 3 weeks	Physical activity level and adherence and IPAQ[f] questionnaire
Sandal et al [45]	selfBack	51 (mean age 45.5, SD 15.0 years)	App (51)	N/A (private)	Improves self-management of low back pain (indirect)	6 weeks	Pain-related disability (RMDQ[g]) and multiple self-reported outcomes

Author	App name	Sample size (age)	Experiment (participant sample size)	Platform[a] (private or public)	App purpose (direct or indirect)	Intervention period	Major outcome indices
Urena et al [33]	m-SFT	7 (53-61 years); the system usability was evaluated by 34 health experts (mean age 36.64 years)	App (7)	Android (private)	Easy-to-use tool for a health practitioner to record and assess the physical condition of older adults (indirect)	N/A	Usability questionnaire
Li et al [49]	Caspar Health App or Website	31 (≥60 years)	App (15); no app (16)	Android and iOS (public)	Postfracture telerehabilitation (direct)	3 weeks	Motor performance, functional performance, and fall efficacy; degree of independence in ADL[h] performance
Kim et al [40]	Fracture Liaison Service	60 (>60 years)	App (60)	Android and iOS (public)	Fall prediction and monitoring (direct)	N/A	Usability
Amorim et al [37]	Fitbit (activity tracker) and IMPACT app	68 (mean age 58.4, SD 13.4 years)	App (34); no app (34)	Android and iOS (public)	Reduces care seeking, pain, and disability in patients with chronic low back pain after treatment discharge (indirect)	15 months	Care seeking, pain levels, and activity limitation
Subasinghe et al [57]	Tap4Bone: MyFitnessPal, Nike Training Club, and QuitBuddy	35 (mean age 23.1 years)	App (18); no app (17)	Android and iOS (public)	MyFitnessPal is a free calorie counter app that helps people track their diet and exercise; Nike Training Club is a free app comprising >100 full-body workouts; QuitBuddy is a smoking cessation internet-based app (indirect)	9 weeks	Feasibility and compliance
Arkkukangas et al [34]	OEP app	12 (70-83 years)	App (12)	N/A (private)	Fall prevention (indirect)	6 weeks	Questionnaire and behavior change
Shebib et al [50]	DCP with sensors	177 (mean age 43, SD 11 years)	App (113); no app (64)	N/A (private)	Aids self-management by engaging patients, and scales personalized therapy for patient-specific needs (indirect)	12 weeks	ODI[i], Korff Pain, and Korff disability
Bedson et al [39]	Keele pain recorder	21 (>18 years)	App (21)	Android (public)	Records pain levels, interference, sleep disturbance, analgesic use, mood, and side effects (indirect)	28 days	Usability and acceptability
Hou et al [55]	eHealth	168 (18-64 years)	app (84); no app (84)	N/A (private)	Telerehabilitation and self-management interventions (indirect)	3, 6, and 12 months	Disease-specific questionnaire (ODI), Visual Analog Scale to record back pain, measures of mental health and life status, which included the EuroQol 5-Dimension health questionnaire
Saran et al [46]	N/A	927 (20-80 years)	App (927)	N/A (private)	Monitors physical activity (indirect)	1 week	Home physical activity
Chhabra et al [54]	Snapcare	93 mean) age 41.4, SD 14.2 years)	App (45); no app (48)	Android (private)	Monitors patient's daily activity levels and symptomatic profile (indirect)	12 weeks	Pain and disability

Author	App name	Sample size (age)	Experiment (participant sample size)	Platform[a] (private or public)	App purpose (direct or indirect)	Intervention period	Major outcome indices
Jakobsen et al [31]	My Osteoporosis Journey	18 (50-65 years)	App (18)	Android and iOS (private)	Provides information and usability questionnaires (direct)	12 weeks	Satisfaction with the app and risk calculation
Lambert et al [56]	PhysiotherapyExercises	80 (34-59 years)	App (40); no app (40)	N/A (private)	Home exercise programs (indirect)	4 weeks	Self-reported exercise adherence, The Patient-Specific Functional Scale, degree of disability, and patient satisfaction with health care service
Rasche et al [43]	Aachen fall prevention app	79 (>50 years)	App (79)	Android and iOS (private)	Self-assessment of older patients at risk for ground-level falls (indirect)	1 year	Objective fall risk and the self-assessed subjective fall risk
Park et al [52]	*Strong bone, Fit body*	82 (<25 years; women)	App (36); no app (38)	Android (private)	Provides feedback and records activity and nutrition (direct)	20 weeks	Bone mineral density, minerals, biochemical markers, food intake diary, knowledge, health belief, and self-efficacy
Tay et al [41]	Calci-app	40 (18-25 years)	App (40)	Android and iOS (private)	Usability questionnaires (direct)	5 days	Dietary calcium intake
Goodman et al [58]	VDC-app	109 (18-25 years)	App (59)	iOS (private)	Provides information and feedback and monitors behavior change (direct)	12 weeks	Vitamin D intake, knowledge, perceptions of vitamin D, blood concentrations of 25(OH)D3
Singler et al [42]	AOTrauma's orthogeriatrics	920 (health professionals)	App (920)	Android and iOS (public)	Delivers the app to surgeons, trainees, and other health care professionals to measure use and evaluate the impact on patient care (direct)	Web-based one-time evaluation	Rating of app and usability

[a]App is available to the public in app stores, or app is not available to the public in app stores.

[b]The study has an indirect relation to osteoporosis.

[c]N/A: not applicable.

[d]SPADE: sleep disturbance, pain, anxiety, depression, and low energy or fatigue.

[e]The study or app has a direct relation to osteoporosis.

[f]IPAQ: International Physical Activity Questionnaire.

[g]RMDQ: Roland-Morris Disability Questionnaire.

[h]ADL: activities of daily living.

[i]ODI: Oswestry Disability Index.

Characteristics of the Included Studies

All selected articles were published in journals over the preceding 8 years (2014-2021), with a notable increase in publications since 2017. The publications comprised feasibility studies [31-39], design and development articles [40-42,61-63], and case studies [43-47]. Among these studies, 41% (13/32) of articles were RCTs [36,37,48-58]. Although most of the articles have a direct relation to osteoporosis [31,40-42,49,52,57,61-63], some of the selected articles refer to apps that are useful and indirectly related to osteoporosis; that is, they are not specifically designed for osteoporosis yet can be potentially useful in managing the disease [32,33,35-37,39,45,51,53-55,58]. The included mHealth apps can be classified into different research themes: (1) monitoring apps (tracking patients' daily nutrition, exercises, and symptoms) [34-36, 38, 40, 41, 44, 46-48, 50-55, 57-59, 64], (2) assessment apps (providing health professionals and patients various tests for assessing patients) [32,33], and (3) measurement apps (measuring certain parameters or variables related to osteoporosis) [43,61-63]. Among all the selected studies, only one of the studies conducted by Ravn Jakobsen et al [65] used participatory design involving all stakeholders, including researchers, women, physicians, health care professionals, and app designers, in the design process of the app named *My Osteoporosis Journey*. After the development stage, they also presented the testing of their collaboratively

designed app with women newly diagnosed with asymptomatic osteoporosis [31].

Characteristics of the Included Apps From Web-Based Market

As of May 2021, we found 33 relevant apps for osteoporosis. Most of the apps identified were found in Google Play (16/33, 48%) and Apple stores (13/33, 39%). Approximately 9% (3/33) of apps were available in the Amazon app store, 3% (1/33) in the GetJar app store, and none in the Galaxy app store.

After removing the overlapping apps across stores, 70% (23/33) of unique apps remained (Figure 3B). Among them, 56% (13/23) were developed to provide educational content on osteoporosis. The educational content covered the diagnosis of the disease, exercises, medications, and diet. It varied among animated videos, recorded videos, short articles, guided audio, expert advice, and graphs.

Table 3 presents all the identified apps in the web-based stores with their main characteristics, including name, operating system, description, users, and classification.

Table 3. Web-based app characteristics.

App name	Operating system	Description	Users	Classification
AACE osteoporosis treatment algorithm[a]	iOS	Provides evidence-based information about the diagnosis, evaluation, and treatment of postmenopausal osteoporosis for endocrinologists, physicians in general, regulatory bodies, health-related organizations, and interested laypersons	Health care professionals	Information and education
Calcium Pro[a]	Android and iOS	Provides information about calcium, parathyroid, osteoporosis, and vitamin D issues; inputs test results for calcium, parathyroid hormone, and vitamin D; analyzes and graphs tests making them easy to understand; tracking tools show calcium and vitamin D levels over time and provide feedback about bone density status; a risk assessment tool for conditions associated with high blood calcium	Patients	Monitoring, education, and assessment
Vitamin-D Pro[a]	iOS	Analyzes and graphs current vitamin D levels, calcium levels, calcium versus parathyroid hormone, bone density, and osteoporosis; teaches how to interpret data and graphs; gives personalized suggestions for next steps; suggests what new blood tests may be necessary; gives topics to discuss with the physician	Patients	Assessment, monitoring, and education tool
Osteoporosis Low Bone Density Weak Bones Diet Help[a]	Android	Provides information about the causes, symptoms, treatment, and the type of diet that one should eat to improve bone density	Patients	Information and education
Bones diseases and treatments[a]	Android	Information about all bone diseases	Patients	Information and education
My Arthritis[a]	Android	Keeps track of symptoms and flares; it can also track diet, exercise, pain, sleep, mood, stress; provides paid training courses with videos, guided audio, and expert advice; sets reminders for appointments and medication; access and share medical records from anywhere; learn about community news, current research, and other information	Patients	Monitoring, assessment, and management
Calcium Calculator[a] (by BC Dairy)	Android	Tool to assess, compare, and plan to introduce enough calcium in daily food	Patients	Monitoring, assessment, and education
Osteoporosis[a] (by AZoMedical)	iOS	Provides regularly updated information and news on osteoporosis	Professionals and patients	News
My Osteoporosis Manager	iOS	Capture detailed information regarding user's health in a digital journal; manage medications and treatments; track osteo-specific symptoms and side effects feedback as easy-to-understand charts that record test results and medication adherence; access patient education materials; share information with a health care provider	Patients	Monitoring, assessment, and management
Osteoporosis (by Focus Media)	Android	Animated videos for learning about osteoporosis disease	Patients	Information and education
Osteoporosis disease	Android	Information about causes, symptoms, treatment, and the type of diet that one should eat to improve bone density	Patients	Information and education
Osteoporosis (by health care tips)	Android	Information and education	Patients	Information and education
Postmenopausal Osteoporosis	Android	Helps in understanding the disease condition through animated videos; it gives an insight into the structure and formation of bones, changes with age, and hormonal levels, particularly during menopause; it also provides information on the onset of osteoporosis, measurement of bone density, treatment, and self-help guidelines	Patients	Information and education
Osteoporosis (by personal remedies)	Android	Comprehensive and actionable nutrition guidelines for how to deal with osteoporosis; recipes, food suggestions, alternative therapies, and remedies	Patients	Information and education
Calcium Supplements	Android	Information about calcium supplements, including who should take them, their health benefits, and potential risks	Patients	Information and education
Osteoporosis AR	Android	Demonstrates a different fictional patient profile using the augmented reality technique that illustrates patient insights, symptoms they are experiencing, and how these agonizing symptoms affect patient's quality of life	Patients	Information and education
Cure for Osteoporosis	Android	Information about raloxifene	Patients	Information

App name	Operating system	Description	Users	Classification
Osteoporosis Risk Calculator	Android	A risk check that calculates whether the user is at risk of fracture or osteoporosis	Patients	Measurement and assessment tool
Hip Fracture Risk Calculator	iOS	Calculates whether the user is at risk of fracture or osteoporosis based on patient demographics	Patients	Measurement and assessment tool
Calcium Calculator	iOS	Calculate calcium intake daily	Patients	Measurement tool
My Osteo-Team	Android and iOS	A social network and support group for those living with osteoporosis; users can acquire practical tips to manage their life with osteoporosis and insights about treatment or therapies	Patients	Social network
Low back pain exercise	Android	Exercises to reduce low back pain	Patients	Information and education
The spine app	Android	Information about back pain	Patients	Information and education
Fracture	Android	Information about fracture prevention	Patients	Information and education

[a]Ranked according to their rating rates, with the highest-ranking rates on the top, and vice versa. The other apps did not have any ratings or reviews. The ranking rate did not reflect the number of times the app was downloaded, and there was no direct relationship between the number of times an app was downloaded and its rating.

Mobile Apps Ranking Based on Features

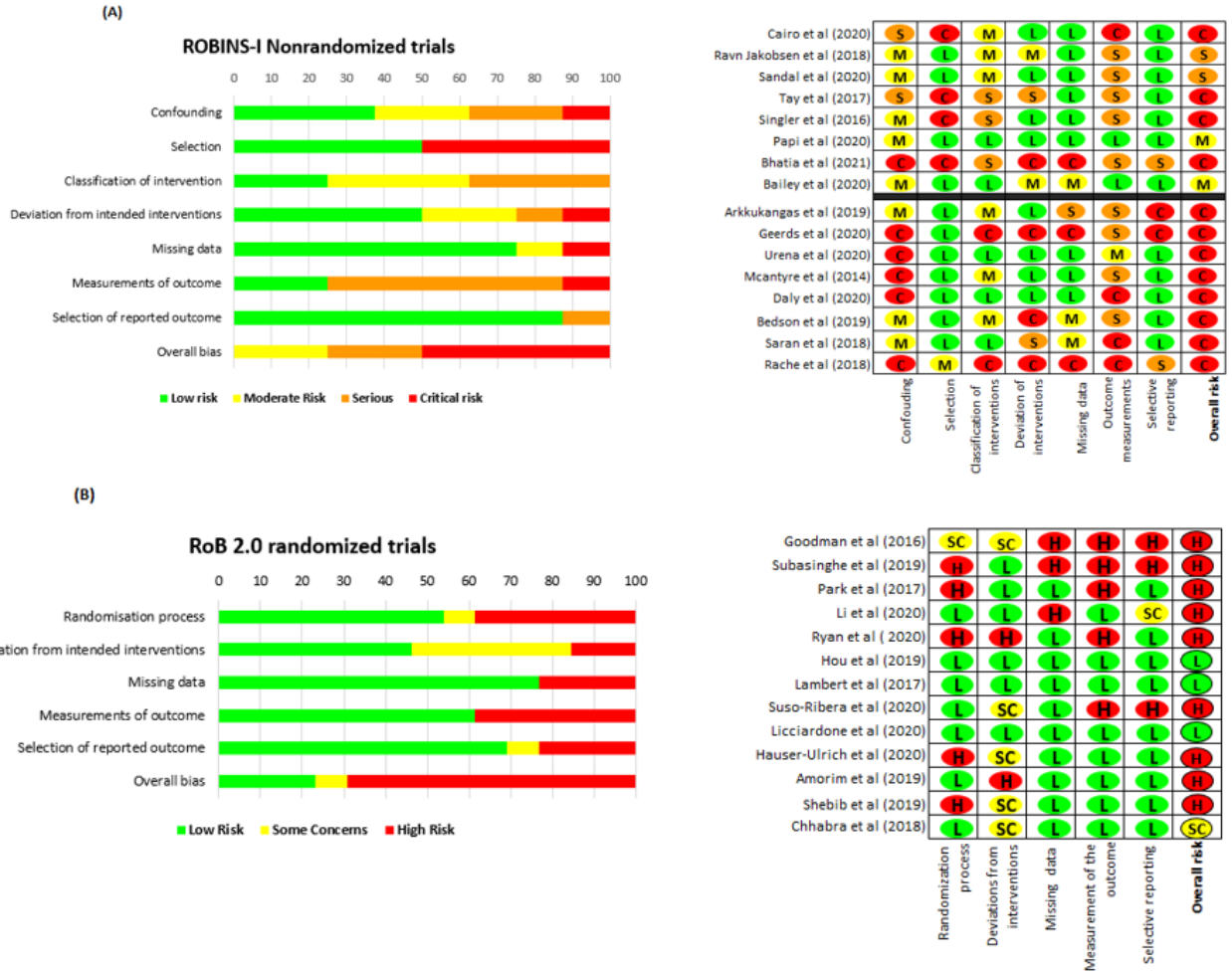
The results of app ranking are presented in Figures 1 and 2. The apps that scored the highest score (equal to or above the raw cutoff, 2.7), both in research and in the web-based market, provided more features, thus, reflecting a more holistic management of osteoporosis and its symptoms [31,40,52,59]. Apps that scored lower had fewer features or were designed for a single purpose, such as measuring spine curvature [63] or BMD [61] (Figure 1A). Approximately 87% (28/32) of the apps provided feedback to users, and 69% (22/32) allowed users to share or export their data and to have an individualized plan based on their individual needs and health goals. Only 3% (1/32) of apps provided a chatbot [53], and 3% (1/32) used AI [45] (Figure 1B).

Similarly, apps in the web-based markets that attained large scores, such as *My Arthritis*, offered more features to assist patients in the management of the disease (Figure 2A). Approximately 75% (17/23) of the web-based apps had good aesthetic and minimalistic designs (simpler designs with the content being the focal points); 70% (16/23) of these apps were designed for diagnosis purposes (Figure 2B).

ROB and Methodological Quality

For the ROB In Nonrandomized Studies of Interventions assessment of nonrandomized clinical trials, 75% (12/16) of studies were at critical ROB, 13% (2/16) at serious risk, and 6% (1/16) at moderate ROB. Among the 13 RCTs assessed using the Cochrane ROB (version 2.0) tool [19], 3 (23%) studies showed a low ROB, and 1 (8%) study exhibited some concerns about the ROB. Approximately 69% (9/13) of RCTs showed a high ROB. Figure 4 summarizes the results of the bias and methodological quality assessments for all studies.

Figure 4. (A) Risk of bias (ROB) assessment for randomized (ROB 2.0) and (B) nonrandomized (ROBIN-I) trials. The studies above the horizontal black line are above the app's cutoff score (2.7) and vice versa [32,33,35-39,41-65]. ROBINS-I: ROB in Nonrandomized Studies of Interventions.

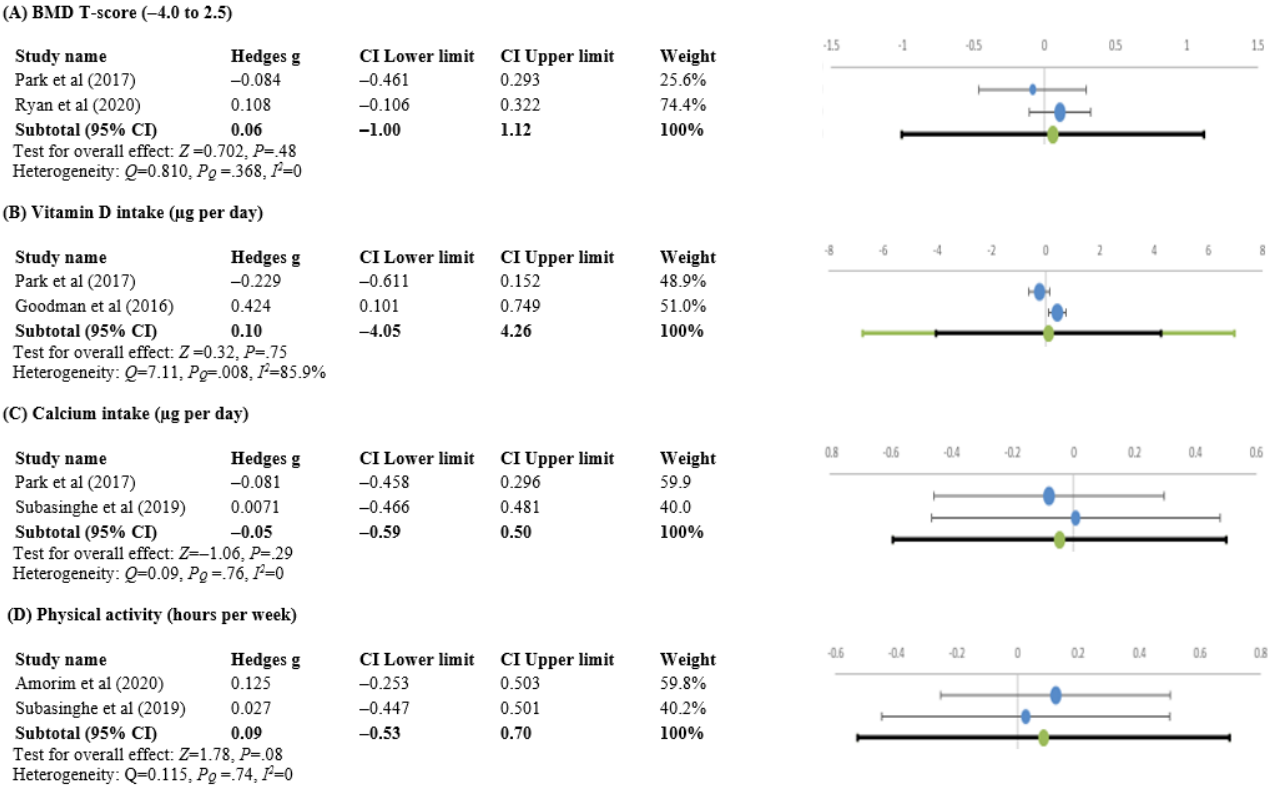


Comparison Between Various Outcomes Before and After App Use

BMD T Score

Approximately 6% (2/32) of studies measured BMD T score at baseline and after 20 weeks [52] and 12 months [48] of using the apps. After initiation, a slight decrease in the mean BMD T score was observed in one of the studies (Hedges *g* –0.084, 95% CI –0.461 to 0.293) [52], and a slight increase was reported in another study (Hedges *g* 0.108, 95% CI –0.106 to 0.322) [48]. The overall change in mean T score was not significant (Hedges *g* 0.06, 95% CI –1.00 to 1.12; Z=0.702; P=.48), with no heterogeneity (Q=0.810; p_Q=0.368; I^2=0; Figure 5A).

Figure 5. Forest plots of Hedges g effect size (95% CI) from individual studies before and after using the app showing changes in (A) bone mineral density (BMD) T score, (B) vitamin D intake (µg per day), (C) calcium intake (µg per day), and (D) physical activity (hours per week) [37,38,52,54-63].



Intake of Vitamin D (µg per Day) and Calcium (mg per Day)

Approximately 6% (2/32) of studies compared the average (µg per day) Vitamin D intake before and after app intervention [52,58]. There was a decrease in intake in one of the studies (Hedges *g* –0.229, 95% CI –0.611 to 0.152) [52], and a moderate increase in intake was observed in another (Hedges *g* 0.424, 95% CI 0.101 to 0.749) [58]. The overall change in intake was not significant (Hedges *g* 0.1, 95% CI –4.05 to 4.26; *Z*=0.32; *P*=.75), with heterogeneity among the studies (*Q*=7.11; p_Q=0.008; I^2=85.9%; Figure 5B).

Approximately 6% (2/32) of studies measured the differences in the average mg per day of calcium intake [52,57]. The daily intake of calcium did not differ significantly before and after app use (*Z*=−1.06; *P*=.29; Hedges *g* −0.05, 95% CI −0.59 to 0.50), with no heterogeneity (*Q*=0.09; p_Q=0.762; I^2=0; Figure 5C).

Physical Activity (Hours per Week)

Approximately 6% (2/32) of studies measured the average number of hours per week of physical activities before and after 15 months [37] or 9 weeks [57] of using the apps. After initiation of the intervention, there was no significant difference observed (*Z*=1.78; *P*=.08; Hedges *g* 0.09, 95% CI −0.53 to 0.70). There was no heterogeneity between the 2 studies (*Q*=0.115; p_Q=0.735; I^2=0; Figure 5D).

Physical Function

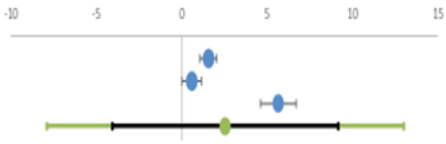
Approximately 6% (2/32) of studies evaluated physical function before and after 4 weeks [56] or 3 weeks [49] of using apps. There was a significant change in physical functioning at the end of the app interventions (Hedges *g* 1.08, 95% CI −5.09 to 7.25; *Z*=2.22; *P*=.03) with heterogeneity (*Q*=7.31; p_Q=0.007; I^2=86.3%; Figure 6A).

Figure 6. Forest plots of Hedges g effect sizes (95% CI) from individual studies before and after using the app showing changes in (A) physical function, (B) well-being, (C) fatigue, and (D) anxiety [37,38,53-55,57-60].

(A) Physical function (0-100)

Study name	Hedges g	CI lower limit	CI upper limit	Weight
Lambert et al (2017)	1.56	1.07	2.05	50.62%
Li et al (2020)	0.59	0.02	1.16	49.38%
Subtotal (95% CI)	**1.08**	**−5.09**	**7.25**	**100%**

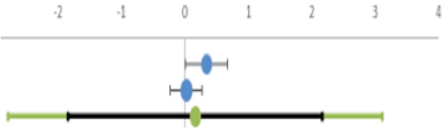
Test for overall effect Z=2.22, P=.03
Heterogeneity: Q=7.31, P_Q=0.007, I^2=86.3%



(B) Well-being (0-100)

Study name	Hedges g	CI lower limit	CI upper limit	Weight
Hauser-Ulrich et al (2019)	0.341	0.009	0.672	44.7
Hou et al (2019)	0.024	−0.229	0.277	55.3
Subtotal (95% CI)	**0.17**	**−1.84**	**2.17**	**100%**

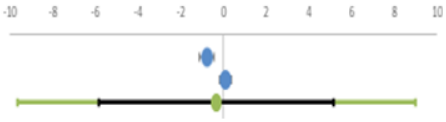
Test for overall effect: Z=1.05, P=.29
Heterogeneity: Q=2.36, P_Q=.12, I^2=57.6%



(C) Fatigue (0-10)

Study name	Hedges g	CI lower limit	CI Upper limit	Weight
Suso-Ribera et al (2020)	−0.775	−1.121	−0.429	49.3
Licciardone et al (2020)	0.091	−0.186	0.309	50.7
Subtotal (95% CI)	**−0.34**	**−5.84**	**5.16**	**100%**

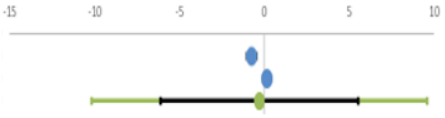
Test for overall effect: Z=−0.776, P=.44
Heterogeneity: Q=15.47, P_Q<.001, I^2=93.5%



(D) Anxiety (0-10)

Study name	Hedges g	CI Lower limit	CI Upper limit	Weight
Suso-Ribera et al (2020)	−0.755	−1.098	−0.411	49.3
Licciardone et al (2020)	0.162	−0.117	0.441	50.7
Subtotal (95% CI)	**−0.29**	**−6.11**	**5.53**	**100%**

Test for overall effect: Z=−0.635, P=.53
Heterogeneity: Q=17.39, P_Q<.001, I^2=94.3%



Well-being

Approximately 6% (2/32) of studies observed changes in well-being from baseline after 12 weeks [53] and 12 months [55] of using the apps. The improvement in well-being was nonsignificant (Hedges g 0.17, 95% CI –0.84 to 2.17; Z=1.05; P=.29), with no heterogeneity (Q=2.36; p_Q=0.125; I^2=57.6%; Figure 6B).

Anxiety and Fatigue

Approximately 6% (2/32) of studies measured changes in anxiety and fatigue at baseline and after 3 months [36] or 4 weeks [51] of intervention. The measured change was not significant for either anxiety (Hedges g –0.29, 95% CI –6.11 to 5.53; Z=–0.635; P=.53), with heterogeneity (Q=17.39; p_Q=0; I^2=94.3%), or fatigue (Hedges g −0.34, 95% CI −5.84 to 5.16), with heterogeneity (Q=15.47; p_Q=0; I^2=93.5%; Figures 6C and 6D).

Pain Intensity

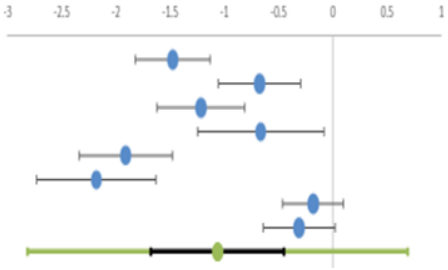
Approximately 25% (8/32) of studies recorded pain intensity before and after initiation of the interventions [36,37,49-51,53-55]. Overall, there was a significant decrease in pain across all studies (Hedges g –1.09, 95% CI –1.68 to –0.45; Z=−4.09; P<.001), with heterogeneity (Q=99.65; p_Q=0; I^2=93%; Figure 7A). A sensitivity analysis was performed to determine whether individual studies had a significant impact on the overall result. No significant differences (P=.81) were observed when excluding individual studies from the analysis (Multimedia Appendix 1, Table S3).

Figure 7. Forest plots of Hedges g effect sizes (95% CI) from individual studies before and after using the app showing changes in (A) pain intensity and (B) disability [37,53,55,57,59,60].

(A) Pain intensity (0-10)

Study name	Hedges g	CI Lower limit	CI Upper limit	Weight
Shebib et al (2019)	−1.474	−1.82	−1.129	12.8%
Amorim et al (2019)	−0.677	−1.056	−0.297	12.7%
Suso-Ribera et al (2020)	−1.216	−1.617	−0.814	12.6%
Li et al (2020)	−0.665	−1.247	−0.082	11.8%
Hou et al (2019)	−1.908	−2.337	−1.48	12.4%
Chhabra et al (2018)	−2.18	−2.729	−1.631	11.8%
Licciardone et al (2020)	−0.183	−0.462	0.097	13.1%
Hauser-Ulrich et al (2019)	−0.311	−0.641	0.019	12.9%
Subtotal (95% CI)	**−1.09**	**−1.68**	**−0.45**	**100%**

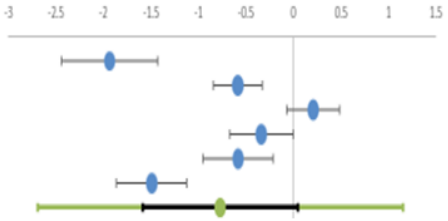
Test for overall effect: Z=−4.09, P <.001
Heterogeneity: Q=99.65, P_Q <.001, I^2=93.0%



(B) Disability (0-100)

Study name	Hedges g	CI Lower limit	CI Upper limit	Weight
Chhabra et al (2018)	−1.936	−2.442	−1.43	15.7%
Shebib et al (2019)	−0.585	−0.843	−0.328	17.2%
Licciardone et al (2020)	0.209	−0.071	0.489	17.1%
Lambert et al (2017)	−0.338	−0.674	−0.002	16.8%
Amorim et al (2019)	−0.584	−0.954	−0.214	16.6%
Hou et al (2019)	−1.491	−1.861	−1.121	16.6%
Subtotal (95% CI)	**−0.77**	**−1.59**	**0.05**	**100%**

Test for overall effect: Z=−2.42, P=.016
Heterogeneity: Q=87.574, P_Q<.001, I^2=94.3%



Disability

Approximately 19% (6/32) of studies evaluated disability [36,37,50,54-56]. The pooled estimate using the random effects model revealed significantly lower levels of disability (Hedges g –0.77, 95% CI –1.59 to 0.05; Z=−2.42; P=.02), with heterogeneity (Q=87.574; p_Q=0; I^2=94.3%; Figure 7B). The sensitivity analysis did not reveal any significant differences (P=.73); Multimedia Appendix 1, Table S4).

Discussion

Principal Findings

The focus of this review was placed on a systematic examination of the available literature on mHealth technologies and apps that can support the self-management of osteoporosis and decision-making for young and older adults. Although some of these apps showed promising results for the use of mHealth technologies in osteoporosis management, there is a lack of evidence in the research to prove the effectiveness of these apps, as validation studies have not been run on all the included apps.

Most (39/52, 75% apps) of the analyzed mHealth apps did not conduct premarket prospective multicenter RCTs. This might be because of the elevated cost of the trials and the long time required to recruit patients [66,67]. In addition, some apps did not publish evidence of their usability and acceptability among users [61,62].

From the scoring system created in Figures 1 and 2, it was possible to observe gaps in the provided features. For instance, only the apps available in the research fields provided feedback to the user, whereas this was not observed in the apps from web-based app stores. This raises an important issue regarding patient accessibility to their data and the overall functionality of these apps.

Our meta-analysis showed that by using the apps, pain scores were significantly reduced in 25% (8/32) of studies [36,37,50,51,53-55,64]. This finding was confirmed by 6% (2/32) of other studies, which found that apps can be beneficial for chronic pain management, especially for patients in an outpatient clinic setting [68,69]. The meta-analysis also showed reduced levels of disability, which is consistent with the findings of Briggs et al [70], who reported reduced disability in patients with osteoarthritis who used digital self-management interventions. In addition, we found that physical function significantly improved after using the apps [56,64].

According to our results, app use had no impact on the physical activity of app users. The meta-analysis also revealed that digital health interventions had no significant impact on the daily intake of calcium and vitamin D or on the BMD trabecular score. It is important to note that patients' adherence to and compliance with the use of mHealth apps are pivotal in ensuring improved health outcomes and successful intervention programs. Some studies reported a high dropout rate in patients who found the intervention boring, time consuming [41,58], or infeasible for daily practice [35]. Another study pointed out that patient attrition led to nonsignificant results at the end of the study [41]. Therefore, any study should ensure to have a comprehensive retention plan for both experimental and control groups.

The data yielded by the meta-analysis demonstrated that using the app had no significant impact on well-being, anxiety, and fatigue scores. This might be explained by the fact that patients self-reported these outcomes in all the evaluated studies without any validation [37,39,44,47,50,54,71], and in many cases, they tended to exaggerate their symptoms in an attempt to prolong the intervention period [50]. To avoid problems arising from self-reporting outcomes, emotionally aware AI techniques could be applied to determine the behavior and emotional state of the user by interpreting their facial expressions while interacting with the app [72] or through emotionally aware chatbots [73].

A review of apps related to osteoporosis in the web-based marketplace resulted in 23 apps. Most of these apps provided information and education, such as disease definition, common symptoms, and suggested exercises to strengthen the bones and enhance physical activity or instructions on healthy nutrition. In addition, none of these available apps addressed the management of the disease after fracture, although fracture is the main complication of osteoporosis. Although osteoporosis is widespread in society, especially among adults, our findings revealed that the number of people who downloaded these osteoporosis-related apps is very limited, as it can barely reach 1000 downloads. Some of these apps did not report any downloads at all. This also indicates that patients or clinicians are hesitant toward the adoption of these new technologies. Unfortunately, these results revealed the poor contribution of research and development toward the field of mHealth apps designed for osteoporosis management and the untrustworthy content that does not have any strong reference [74].

Information privacy is an important issue in mHealth apps because of the sensitive nature of information gathered from users [75]. All identified apps in the web-based app stores were free, with pop-up advertisements every once in a while. Apparently, this creates a distraction for the user and sets some doubts about the way the collected data are used (eg, to create targeted advertisements according to the user's profile). Apparently, this needs to be consented to by the user, following data privacy and security protocols, such as the General Data Protection Regulation [76].

Our findings show that only one of the identified studies [31] used a participatory design process to develop their app. Cooperative or participatory design involves stakeholders, designers, researchers, and end users in the early stages of the design process to ensure that the developed app meets the proper needs of its intended end users [77]. This entails that all stakeholders have equal input in the interaction design, which will nurture a more creative development atmosphere. Moreover, this cocreation session gives stakeholders a sense of ownership of the ideas that allow them to comprehend the thinking behind design decisions and improve their satisfaction levels. The involvement of health professionals in the design will also prevent any safety risks arising from inaccurate or unreliable digital tools [78].

Overcoming Challenges in Osteoporosis mHealth Apps

Overview

This review shows that mHealth apps that use self-management support principles in primary care have the potential to have a positive effect on the management of chronic diseases. However, there is reluctance in the adoption of these digital technologies in health care. The main obstacles delaying the integration of these technological tools in osteoporosis care could be summarized as (1) weak or no involvement of health care professionals in the design process, (2) reluctance of clinicians who believe that mHealth apps might replace them, (3) lack of reliable tools and strict regulations, (4) privacy and security concerns, (5) data availability and visualization, (6) inconsistent data collection standards, (7) difficulties in acquiring and analyzing data, and (8) low retention rates of participants (Table 4).

Table 4. DRsa and CRsb for overcoming the identified limitations or barriers in digital health technologies for osteoporosis.

Identified limitation or barrier and related aspect	Recommendation
App designers and developers without supporting information from clinicians, resulting in a very technologically focused and problem-oriented approach in the design of mHealth apps	
Design perspective	• DR1: involve all the stakeholders in all the stages of user requirements, design, and development using a participatory design approach (cocreation)
Clinical perspective	• CR1: active participation in the design, development, and testing stages
Clinicians' reluctance in adopting mHealth [c] apps as they envisage that they will replace them	
Clinical perspective	• CR1: adopt mHealth technologies in daily practices and in clinical care (measurement, assessment, and recording data) • CR2: recommend trustworthy apps to their patients • CR3: use mHealth apps to effectively communicate with patients and other health care professionals through the integration of wearables and IoTd
Lack of trustworthy and available smart tools and strict regulations on mHealth tools	
Design perspective	• DR1: use adaptive learning algorithms (eg, AIe and machine or deep learning) in the app to make more personalized recommendations and treatments • DR2: incorporate clinically validated monitoring, measurement, and assessment tools in the designed app
Clinical perspective	• CR1: evaluate mHealth measurement and assessment tools by concerned clinical experts before disseminating them to public
Underestimation of the security risk and the elevated cost of implementing strong data security and privacy rules	
Design perspective	• DR1: implement stringent security regulations (eg, GDPRf [79]) to protect users' information from any data penetration (security and privacy by design)
Available data are provider oriented rather than patient accessible; limited existing guidelines on how to optimize user interfaces for patients, providers, or both	
Design perspective	• DR1: allow patients to access their data (GDPR enforcement in design) • DR2: generate feedback and plans (for diet and exercises) based on the gathered data to keep patients engaged and motivated
Inconsistent data collection standards, complexity of data, and lack of quality assurance processes (data cannot be verified)	
Design perspective	• DR1: use passive and active gathering of data (medication, symptoms, nutrition management, and physical exercising), in addition to the data gathered from any wearables or IoT sensors
Clinical perspective	• CR1: combine conventional clinical assessment with the app assessment
Difficulties in acquiring, analyzing, and applying structured and unstructured data to treat or manage diseases	
Design perspective	• DR1: apply AI-based techniques that help with the prediction, diagnosis, and treatment or management of diseases
Low retention rates of participants	
Design perspective	• DR1: provide valuable feedback to the user • DR2: use simple and straightforward interfaces • DR3: continuously update users' data • DR4: offer financial incentives for healthy habit changes

[a]DR: design-related recommendation.

[b]CR: clinical recommendation.

[c]mHealth: mobile health.

[d]IoT: Internet of Things.

[e]AI: artificial intelligence.

[f]GDPR: General Data Protection Regulation.

To overcome these obstacles, we propose design-related and clinical recommendations for mHealth apps to support patients at risk of or diagnosed with osteoporosis in self-management and involve them in decision-making regarding treatment and

intervention options with clinicians. These guidelines are not only limited to apps targeting osteoporosis self-management but can also be applied to any chronic disease self-management app.

Design-Related Recommendations

Co-design

Before creating an app, we should emphasize the role of end users, including patients and health professionals, in the development process. Users should be involved at various points and levels in the design process to improve their understanding of their needs, requirements, interactions, and appreciations before, during, and after developing the app. This co-design will ensure that the developed app meets end user purposes, leading to more effective results [80].

Integration of AI and Machine Learning in Data Acquisition and in Decision-making

An enhanced and intelligent version of the mHealth app can perform long-term management of osteoporosis through internet-based coaching using AI and big data analysis. In addition to health care professionals, AI can play an important role in the decision-making process and in the entire self-management process of osteoporosis. Conventional systems used for processing health data are less accurate and lack convergence compared with AI-supported systems [81]. Machine learning methods, more specifically adaptive learning algorithms, integrated into mHealth apps will make them tailored to an individual's behavior and characteristics, thus improving the effectiveness of the intervention [82]. Such smart mHealth apps could unobtrusively acquire and effectively analyze sensorial and behavioral cloud-archived big data from adults' interactions with smart devices (smartphones, smartwatches, and Internet of Things) in their daily living environment [83].

Owing to the significant advances and progress in AI in the past few years, chatbots have been gaining momentum in the eHealth world. Therefore, we believe that a bot framework can be incorporated with virtual reality technology and low-cost Internet of Things to create a user engagement schema for long-term monitoring of osteoporosis, where the patient will be active and maintain an improved quality of life.

Envisioning of a Smart Tool With User-Centered Orientation in Osteoporosis Management

An innovative technological tool (mHealth app) should offer an integrated platform for informed healthy living indoors or outdoors to assist patients with osteoporosis (or at risk) in different aspects of life, including physical activity, nutrition, medication intake, fall prevention, emotional wellness, and socialization. The design of such tools could include monitoring, combining both passive (via the interaction with smartphone or smartwatch or wearables) and active gathering of data (eg, about medication, nutrition management, and physical exercising). Then, on the basis of the gathered data, AI-driven data analysis processes could be involved in providing personalized feedback to the patient and informing the related physician, guiding personalized recommendations and interventions for osteoporosis risk assessment [84]. In this way, the patients will be kept aware of their progress in osteoporosis self-management over a certain period, notified in case of any increased risk [85], motivated and engaged in using the app, and follow the personalized intervention program. The mHealth app should provide the user with various ways of data visualization and access at any time, scaffolding a participatory management of the disease.

Enhanced Security and Privacy Measures

mHealth app developers must ensure that collected user data are secured to maintain the integrity, availability, confidentiality, and resilience of the data [86]. Security procedures should comply with the best practices and regulations (eg, General Data Protection Regulation [79]). Users should be aware of the techniques used to safeguard their personal information and the authentication methods used. These enhanced security measures will make it possible to leverage mHealth tools in daily practice for both clinicians and patients.

Improving Participants' Retention Rates

The success and effectiveness of any mHealth app intervention are directly related to user retention [87]. Therefore, to attain the maximum clinical benefit from the app, designers should ensure that users adhere for the long term to mHealth apps [88]. Various plans could be adopted by designers to re-engage and retain users. mHealth apps should be designed with simple and easy-to-use interfaces as many users refrain from using mHealth apps because of their complicated implementation [89]. Another approach is to continuously notify users about their progress and provide them with positive feedback. An app publisher could also provide users with financial incentives or awards if they achieve a certain healthy goal; for instance, this incentive could be free health insurance or a free subscription to the nearest gym. Applying these design techniques will retain a larger number of participants, resulting in a better impact of the mHealth app interventions.

Clinical Recommendations

Adoption of Digital Therapeutics

As patients increasingly turn to mHealth apps and devices, clinicians must consider the value of these apps and embrace them to deliver enhanced care. They should adopt more mHealth technologies in their daily practices or workflows and integrate the data into electronic medical records. However, physicians should refrain from recommending apps that have been created without the involvement of medical experts or appropriate testing validation, especially if claims made by app developers are fraudulent. To ensure that their requirements are met and to deliver better outcomes, physicians should actively participate in the design, development, and testing of these mHealth apps [90]. In validated (Food and Drug Administration–approved and *Conformité Européenne*–marked) cases only, *prescription* of these apps can be envisioned, as a form of digital therapeutics, along with wearable devices to allow remote and real-time health monitoring and care delivery [91].

mHealth Apps for Communication and Continuous Improvement of Health Care

mHealth apps create a sense of partnership between patients and health care professionals by allowing patients to play a

more active role in their health care. Moreover, digital health will improve patients' engagement with their treatment, something that physicians have previously struggled to do between visits [75]. They also allow proper communication between physicians, patients, and other health care professionals [92]. These tools, assisted by AI and machine learning, represent a rich source of data for clinicians that can be used in medical research to continuously improve the overall delivery of care. It is important to note that these apps are not designed to replace clinicians; on the contrary, they support their decision-making process and workflow. Physicians should not consider these apps as opponents or competitors but rather as an opportunity to enable a streamlined high-quality health care delivery process by capturing and analyzing more data, reaching and monitoring a larger number of patients remotely, and perpetually advancing their clinical practices.

Limitations

Despite this in-depth analysis, some limitations can be identified in this review. In particular, we refrained from excluding studies based on certain quality criteria, such as study design or sample size, which resulted in large variations in the measurements of outcomes. Moreover, there is a lack of apps that only target osteoporosis; therefore, we included apps that we thought were *useful* for osteoporosis. As no articles in any other language were identified as eligible, the risk of language bias in our selection was negligible. Regarding other biases in the selection, we were cautious in our selection by selecting all related articles in the fields regardless of their outcomes or study design. Finally, some feasibility and development studies were included with the intention of understanding any novel approaches being tested or developed. Such studies seem promising to achieve potential outcomes; however, as these apps were either not tested or tested but with a relatively small sample size, it is difficult to determine whether such solutions can be adopted in the mainstream.

Conclusions

Given the identified lack of effective mHealth apps with a holistic approach to osteoporosis self-management, this review holds the potential to bridge this gap by proposing a technological tool that goes beyond apps that simply provide information about osteoporosis and creates an individualized care management plan that goes beyond clinical measures. The latter perspective extends the view of mHealth apps from the initial focus on promoting specific behavior, such as healthy nutrition, physical activity, or adherence to medications, to patients' engagement and empowerment. Moreover, it strengthens collaboration between patients and caregivers by not limiting it to health institutions. In view of the vast quantity of mHealth apps available, it is important for app developers and researchers to identify the proper needs of patients with osteoporosis, adopting a cocreation strategy to create more patient-centered and effective disease management solutions.

Acknowledgments

This work was supported by Khalifa University of Science and Technology under award number CIRA-2020-031.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist, detailed search terms, and sensitivity analysis.
[DOCX File , 24 KB - mhealth_v10i4e32557_app1.docx]

References

1. Sözen T, Özışık L, Başaran N. An overview and management of osteoporosis. Eur J Rheumatol 2017;4(1):46-56 [FREE Full text] [doi: 10.5152/eurjrheum.2016.048] [Medline: 28293453]
2. Akkawi I, Zmerly H. Osteoporosis: current concepts. Joints 2018;6(2):122-127 [FREE Full text] [doi: 10.1055/s-0038-1660790] [Medline: 30051110]
3. Johnell O, Kanis JA. An estimate of the worldwide prevalence and disability associated with osteoporotic fractures. Osteoporos Int 2006;17(12):1726-1733. [doi: 10.1007/s00198-006-0172-4] [Medline: 16983459]
4. Ballane G, Cauley JA, Luckey MM, El-Hajj Fuleihan G. Worldwide prevalence and incidence of osteoporotic vertebral fractures. Osteoporos Int 2017;28(5):1531-1542. [doi: 10.1007/s00198-017-3909-3] [Medline: 28168409]
5. Reginster JY, Burlet N. Osteoporosis: a still increasing prevalence. Bone 2006;38(2 Suppl 1):S4-S9. [doi: 10.1016/j.bone.2005.11.024] [Medline: 16455317]
6. Kelley GA, Kelley KS. Dropouts and compliance in exercise interventions targeting bone mineral density in adults: a meta-analysis of randomized controlled trials. J Osteoporos 2013;2013:250423 [FREE Full text] [doi: 10.1155/2013/250423] [Medline: 23862095]
7. Kothawala P, Badamgarav E, Ryu S, Miller RM, Halbert RJ. Systematic review and meta-analysis of real-world adherence to drug therapy for osteoporosis. Mayo Clin Proc 2007;82(12):1493-1501. [doi: 10.1016/S0025-6196(11)61093-8] [Medline: 18053457]

8. Agnihothri S, Cui L, Delasay M, Rajan B. The value of mHealth for managing chronic conditions. Health Care Manag Sci 2020;23(2):185-202. [doi: 10.1007/s10729-018-9458-2] [Medline: 30382448]
9. Rotheram-Borus MJ, Ingram BL, Swendeman D, Lee A. Adoption of self-management interventions for prevention and care. Prim Care 2012;39(4):649-660 [FREE Full text] [doi: 10.1016/j.pop.2012.08.006] [Medline: 23148958]
10. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. PLoS One 2017;12(2):e0170581 [FREE Full text] [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]
11. Bradway M, Carrion C, Vallespin B, Saadatfard O, Puigdomènech E, Espallargues M, et al. mHealth assessment: conceptualization of a global framework. JMIR Mhealth Uhealth 2017;5(5):e60 [FREE Full text] [doi: 10.2196/mhealth.7291] [Medline: 28465282]
12. Koetsier J. There are now 8.9 million mobile apps, and China is 40% of mobile app spending. Forbes. 2020. URL: https://tinyurl.com/2p9447by [accessed 2021-05-12]
13. mHealth new horizons for health through mobile technologies: based on the findings of the second global survey on eHealth: global observatory for eHealth series - volume 3. World Health Organization Global Observatory for eHealth. 2011. URL: https://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2021-05-12]
14. Pérez Sust P, Solans O, Fajardo JC, Medina Peralta M, Rodenas P, Gabaldà J, et al. Turning the crisis into an opportunity: digital health strategies deployed during the COVID-19 outbreak. JMIR Public Health Surveill 2020;6(2):e19106 [FREE Full text] [doi: 10.2196/19106] [Medline: 32339998]
15. Latif S, Rana R, Qadir J, Ali A, Imran MA, Younis MS. Mobile health in the developing world: review of literature and lessons from a case study. IEEE Access 2017;5:11540-11556. [doi: 10.1109/access.2017.2710800]
16. Barbabella F, Melchiorre MG, Quattrini S, Papa R, Lamura G. In: Richardson E, van Ginneken E, editors. How can eHealth improve care for people with multimorbidity in Europe?. Copenhagen, Denmark: European Observatory on Health Systems and Policies; 2017.
17. Tsamlag L, Wang H, Shen Q, Shi Y, Zhang S, Chang R, et al. Applying the information-motivation-behavioral model to explore the influencing factors of self-management behavior among osteoporosis patients. BMC Public Health 2020;20(1):198 [FREE Full text] [doi: 10.1186/s12889-020-8292-x] [Medline: 32028930]
18. Van de Velde D, De Zutter F, Satink T, Costa U, Janquart S, Senn D, et al. Delineating the concept of self-management in chronic conditions: a concept analysis. BMJ Open 2019;9(7):e027775 [FREE Full text] [doi: 10.1136/bmjopen-2018-027775] [Medline: 31315862]
19. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
20. Castro VM, McCoy TH, Cagan A, Rosenfield HR, Murphy SN, Churchill SE, et al. Stratification of risk for hospital admissions for injury related to fall: cohort study. BMJ 2014;349:g5863 [FREE Full text] [doi: 10.1136/bmj.g5863] [Medline: 25954985]
21. Suurmond R, van Rhee H, Hak T. Introduction, comparison, and validation of meta-essentials: a free and simple tool for meta-analysis. Res Synth Methods 2017;8(4):537-553 [FREE Full text] [doi: 10.1002/jrsm.1260] [Medline: 28801932]
22. Wasserman S. J Educ Stat 1988;13(1):75-78. [doi: 10.2307/1164953]
23. Cochran WG. The comparison of percentages in matched samples. Biometrika 1950;37(3/4):256-266. [doi: 10.2307/2332378]
24. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21(11):1539-1558. [doi: 10.1002/sim.1186] [Medline: 12111919]
25. Venkatesh V, Thong JY, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157-178 [FREE Full text] [doi: 10.2307/41410412]
26. Nielsen J, Mack RL. Usability inspection methods. New York, NY: John Wiley & Sons; 1994.
27. Nielsen J, Molich R. Heuristic evaluation of user interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1990 Presented at: CHI '90; April 1-5, 1990; Seattle, WA p. 249-256. [doi: 10.1145/97243.97281]
28. Handel MJ. mHealth (mobile health)-using Apps for health and wellness. Explore (NY) 2011;7(4):256-261. [doi: 10.1016/j.explore.2011.04.011] [Medline: 21724160]
29. Li J, Zhang C, Li X, Zhang C. Patients' emotional bonding with MHealth apps: an attachment perspective on patients' use of MHealth applications. Int J Inf Manage 2020;51(C):102054. [doi: 10.1016/j.ijinfomgt.2019.102054]
30. Mendiola MF, Kalnicki M, Lindenauer S. Valuable features in mobile health apps for patients and consumers: content analysis of apps and user ratings. JMIR Mhealth Uhealth 2015;3(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4283] [Medline: 25972309]
31. Ravn Jakobsen P, Hermann AP, Søndergaard J, Wiil UK, Clemensen J. Help at hand: women's experiences of using a mobile health application upon diagnosis of asymptomatic osteoporosis. SAGE Open Med 2018;6:2050312118807617 [FREE Full text] [doi: 10.1177/2050312118807617] [Medline: 30397471]
32. Papi E, Chiou SY, McGregor AH. Feasibility and acceptability study on the use of a smartphone application to facilitate balance training in the ageing population. BMJ Open 2020;10(12):e039054 [FREE Full text] [doi: 10.1136/bmjopen-2020-039054] [Medline: 33268409]

33. Ureña R, Chiclana F, Gonzalez-Alvarez A, Herrera-Viedma E, Moral-Munoz JA. m-SFT: a novel mobile health system to assess the elderly physical condition. Sensors (Basel) 2020;20(5):1462 [FREE Full text] [doi: 10.3390/s20051462] [Medline: 32155931]
34. Arkkukangas M, Cederbom S, Tonkonogi M, Umb Carlsson Õ. Older adults' experiences with mHealth for fall prevention exercise: usability and promotion of behavior change strategies. Physiother Theory Pract 2021;37(12):1346-1352. [doi: 10.1080/09593985.2020.1712753] [Medline: 31910707]
35. Geerds MA, Nijmeijer WS, Hegeman JH, Vollenbroek-Hutten MM. Mobile app for monitoring 3-month postoperative functional outcome after hip fracture: usability study. JMIR Hum Factors 2020;7(3):e16989 [FREE Full text] [doi: 10.2196/16989] [Medline: 32924949]
36. Licciardone JC, Pandya V. Feasibility trial of an eHealth intervention for health-related quality of life: implications for managing patients with chronic pain during the COVID-19 pandemic. Healthcare (Basel) 2020;8(4):381 [FREE Full text] [doi: 10.3390/healthcare8040381] [Medline: 33019676]
37. Amorim AB, Pappas E, Simic M, Ferreira ML, Jennings M, Tiedemann A, et al. Integrating mobile-health, health coaching, and physical activity to reduce the burden of chronic low back pain trial (IMPACT): a pilot randomised controlled trial. BMC Musculoskelet Disord 2019;20(1):71 [FREE Full text] [doi: 10.1186/s12891-019-2454-y] [Medline: 30744606]
38. Daly RM, Gianoudis J, Hall T, Mundell NL, Maddison R. Feasibility, usability, and enjoyment of a home-based exercise program delivered via an exercise app for musculoskeletal health in community-dwelling older adults: short-term prospective pilot study. JMIR Mhealth Uhealth 2021;9(1):e21094 [FREE Full text] [doi: 10.2196/21094] [Medline: 33439147]
39. Bedson J, Hill J, White D, Chen Y, Wathall S, Dent S, et al. Development and validation of a pain monitoring app for patients with musculoskeletal conditions (The Keele pain recorder feasibility study). BMC Med Inform Decis Mak 2019;19(1):24 [FREE Full text] [doi: 10.1186/s12911-019-0741-z] [Medline: 30683106]
40. Kim SW, Won YJ, Chae DS, Chang HJ. A new fracture liaison service using the mobile application and IoT sensor. Annu Int Conf IEEE Eng Med Biol Soc 2019;2019:3486-3489. [doi: 10.1109/EMBC.2019.8857094] [Medline: 31946629]
41. Tay I, Garland S, Gorelik A, Wark JD. Development and testing of a mobile phone app for self-monitoring of calcium intake in young women. JMIR Mhealth Uhealth 2017;5(3):e27 [FREE Full text] [doi: 10.2196/mhealth.5717] [Medline: 28270379]
42. Singler K, Roth T, Beck S, Cunningham M, Gosch M. Development and initial evaluation of a point-of-care educational app on medical topics in orthogeriatrics. Arch Orthop Trauma Surg 2016;136(1):65-73 [FREE Full text] [doi: 10.1007/s00402-015-2366-8] [Medline: 26646844]
43. Rasche P, Mertens A, Bröhl C, Theis S, Seinsch T, Wille M, et al. The "Aachen fall prevention App" - a smartphone application app for the self-assessment of elderly patients at risk for ground level falls. Patient Saf Surg 2017;11:14 [FREE Full text] [doi: 10.1186/s13037-017-0130-4] [Medline: 28503199]
44. Bhatia A, Kara J, Janmohamed T, Prabhu A, Lebovic G, Katz J, et al. User engagement and clinical impact of the manage my pain app in patients with chronic pain: a real-world, multi-site trial. JMIR Mhealth Uhealth 2021;9(3):e26528 [FREE Full text] [doi: 10.2196/26528] [Medline: 33661130]
45. Sandal LF, Øverås CK, Nordstoga AL, Wood K, Bach K, Hartvigsen J, et al. A digital decision support system (selfBACK) for improved self-management of low back pain: a pilot study with 6-week follow-up. Pilot Feasibility Stud 2020;6:72 [FREE Full text] [doi: 10.1186/s40814-020-00604-2] [Medline: 32489674]
46. Saran T, Pedrycz A, Mucha D, Mucha D. Follow-up monitoring of physical activity after rehabilitation by means of a mobile application: effectiveness of measurements in different age groups. Adv Clin Exp Med 2018;27(8):1037-1044 [FREE Full text] [doi: 10.17219/acem/69131] [Medline: 29962117]
47. Bailey JF, Agarwal V, Zheng P, Smuck M, Fredericson M, Kennedy DJ, et al. Digital care for chronic musculoskeletal pain: 10,000 participant longitudinal cohort study. J Med Internet Res 2020;22(5):e18250 [FREE Full text] [doi: 10.2196/18250] [Medline: 32208358]
48. Ryan P, Brown RL, Csuka ME, Papanek P. Efficacy of osteoporosis prevention smartphone app. Nurs Res 2020;69(1):31-41 [FREE Full text] [doi: 10.1097/NNR.0000000000000392] [Medline: 31568199]
49. Li CT, Hung GK, Fong KN, Gonzalez PC, Wah S, Tsang HW. Effects of a home-based occupational therapy telerehabilitation via smartphone for outpatients after hip fracture surgery: a feasibility randomised controlled study. J Telemed Telecare 2020:1357633X20932434. [doi: 10.1177/1357633X20932434] [Medline: 32594832]
50. Shebib R, Bailey JF, Smittenaar P, Perez DA, Mecklenburg G, Hunter S. Randomized controlled trial of a 12-week digital care program in improving low back pain. NPJ Digit Med 2019;2:1 [FREE Full text] [doi: 10.1038/s41746-018-0076-7] [Medline: 31304351]
51. Suso-Ribera C, Castilla D, Zaragozá I, Mesas Á, Server A, Medel J, et al. Telemonitoring in chronic pain management using smartphone apps: a randomized controlled trial comparing usual assessment against app-based monitoring with and without clinical alarms. Int J Environ Res Public Health 2020;17(18):6568 [FREE Full text] [doi: 10.3390/ijerph17186568] [Medline: 32916983]
52. Park YJ, Lee SJ, Shin NM, Shin H, Jeon S, Lee J, et al. Application and effect of mobiletype-bone health intervention in Korean young adult women with low bone mass: a randomized control trial. Asian Nurs Res (Korean Soc Nurs Sci) 2017;11(1):56-64 [FREE Full text] [doi: 10.1016/j.anr.2017.03.005] [Medline: 28388981]

53. Hauser-Ulrich S, Künzli H, Meier-Peterhans D, Kowatsch T. A smartphone-based health care chatbot to promote self-management of chronic pain (SELMA): pilot randomized controlled trial. JMIR Mhealth Uhealth 2020;8(4):e15806 [FREE Full text] [doi: 10.2196/15806] [Medline: 32242820]
54. Chhabra HS, Sharma S, Verma S. Smartphone app in self-management of chronic low back pain: a randomized controlled trial. Eur Spine J 2018;27(11):2862-2874. [doi: 10.1007/s00586-018-5788-5] [Medline: 30324496]
55. Hou J, Yang R, Yang Y, Tang Y, Deng H, Chen Z, et al. The effectiveness and safety of utilizing mobile phone-based programs for rehabilitation after lumbar spinal surgery: multicenter, prospective randomized controlled trial. JMIR Mhealth Uhealth 2019;7(2):e10201 [FREE Full text] [doi: 10.2196/10201] [Medline: 30785406]
56. Lambert TE, Harvey LA, Avdalis C, Chen LW, Jeyalingam S, Pratt CA, et al. An app with remote support achieves better adherence to home exercise programs than paper handouts in people with musculoskeletal conditions: a randomised trial. J Physiother 2017;63(3):161-167 [FREE Full text] [doi: 10.1016/j.jphys.2017.05.015] [Medline: 28662834]
57. Subasinghe AK, Garland SM, Gorelik A, Tay I, Wark JD. Using mobile technology to improve bone-related lifestyle risk factors in young women with low bone mineral density: feasibility randomized controlled trial. JMIR Form Res 2019;3(1):e9435 [FREE Full text] [doi: 10.2196/formative.9435] [Medline: 30801253]
58. Goodman S, Morrongiello B, Meckling K. A randomized, controlled trial evaluating the efficacy of an online intervention targeting vitamin D intake, knowledge and status among young adults. Int J Behav Nutr Phys Act 2016;13(1):116 [FREE Full text] [doi: 10.1186/s12966-016-0443-1] [Medline: 27836017]
59. Cairo J, Williams L, Bray L, Goetzke K, Perez AC. Evaluation of a mobile health intervention to improve wellness outcomes for breast cancer survivors. J Patient Cent Res Rev 2020;7(4):313-322 [FREE Full text] [doi: 10.17294/2330-0698.1733] [Medline: 33163551]
60. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
61. Sulla-Torres J, Gutierrez-Quintanilla A, Pinto-Rodriguez H, Gomez-Campos R, Cossio-Bolanos M. Quality in use of an android-based mobile application for calculation of bone mineral density with the standard ISO/IEC 25022. Int J Adv Comput Sci Appl 2020;11(8):158-163. [doi: 10.14569/ijacsa.2020.0110821]
62. Bernardi HL, da Motta LB. Development of an app as a tool to support research and the prevention of osteoporosis. Rev Bras Geriatr Gerontol 2018;21(4):408-418. [doi: 10.1590/1981-22562018021.170189]
63. MacIntyre NJ, Recknor CP, Recknor JC. Spine curve measures taken in individuals with osteoporosis and osteopenia using the IONmed mobile phone application are highly consistent with digital inclinometer measures. Int Sch Res Notices 2014;2014:1-6. [doi: 10.1155/2014/508414]
64. Li J, Hodgson N, Lyons MM, Chen KC, Yu F, Gooneratne NS. A personalized behavioral intervention implementing mHealth technologies for older adults: a pilot feasibility study. Geriatr Nurs 2020;41(3):313-319 [FREE Full text] [doi: 10.1016/j.gerinurse.2019.11.011] [Medline: 31810730]
65. Ravn Jakobsen P, Hermann AP, Søndergaard J, Wiil UK, Clemensen J. Development of an mHealth application for women newly diagnosed with osteoporosis without preceding fractures: a participatory design approach. Int J Environ Res Public Health 2018;15(2):330 [FREE Full text] [doi: 10.3390/ijerph15020330] [Medline: 29438343]
66. Pham Q, Wiljer D, Cafazzo JA. Beyond the randomized controlled trial: a review of alternatives in mHealth clinical trial methods. JMIR Mhealth Uhealth 2016;4(3):e107 [FREE Full text] [doi: 10.2196/mhealth.5720] [Medline: 27613084]
67. Tarricone R, Petracca F, Ciani O, Cucciniello M. Distinguishing features in the assessment of mHealth apps. Expert Rev Pharmacoecon Outcomes Res 2021;21(4):521-526. [doi: 10.1080/14737167.2021.1891883] [Medline: 33663324]
68. Thurnheer SE, Gravestock I, Pichierri G, Steurer J, Burgstaller JM. Benefits of mobile apps in pain management: systematic review. JMIR Mhealth Uhealth 2018;6(10):e11231 [FREE Full text] [doi: 10.2196/11231] [Medline: 30348633]
69. Pfeifer AC, Uddin R, Schröder-Pfeifer P, Holl F, Swoboda W, Schiltenwolf M. Mobile application-based interventions for chronic pain patients: a systematic review and meta-analysis of effectiveness. J Clin Med 2020;9(11):3557 [FREE Full text] [doi: 10.3390/jcm9113557] [Medline: 33167300]
70. Safari R, Jackson J, Sheffield D. Digital self-management interventions for people with osteoarthritis: systematic review with meta-analysis. J Med Internet Res 2020;22(7):e15365 [FREE Full text] [doi: 10.2196/15365] [Medline: 32706657]
71. Sandal LF, Stochkendahl MJ, Svendsen MJ, Wood K, Øverås CK, Nordstoga AL, et al. An app-delivered self-management program for people with low back pain: protocol for the selfBACK randomized controlled trial. JMIR Res Protoc 2019;8(12):e14720 [FREE Full text] [doi: 10.2196/14720] [Medline: 31793897]
72. Emotion-aware intelligent systems for sustainable healthcare practices. Engineering in Medicine & Biology Society. URL: https://www.embs.org/emotion-aware-intelligent-systems-for-sustainable-healthcare-practices/ [accessed 2021-11-22]
73. Ghandeharioun A, Mcduff D, Czerwinski M, Rowan K. EMMA: an emotion-aware wellbeing chatbot. In: 8th International Conference on Affective Computing and Intelligent Interaction. 2019 Presented at: ACII '19; September 3-6, 2019; Cambridge, UK p. 1-7. [doi: 10.1109/acii.2019.8925455]
74. van Haasteren A, Gille F, Fadda M, Vayena E. Development of the mHealth app trustworthiness checklist. Digit Health 2019;5:2055207619886463 [FREE Full text] [doi: 10.1177/2055207619886463] [Medline: 31803490]

75. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med 2020;3:4 [FREE Full text] [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]
76. Out of control: how consumers are exploited by the online advertising industry. ForbrukerRådet. 2020. URL: https://fil.forbrukerradet.no/wp-content/uploads/2020/01/2020-01-14-out-of-control-final-version.pdf [accessed 2021-11-29]
77. Scariot CA, Heemann A, Padovani S. Understanding the collaborative-participatory design. Work 2012;41 Suppl 1:2701-2705. [doi: 10.3233/WOR-2012-0656-2701] [Medline: 22317129]
78. Akbar S, Coiera E, Magrabi F. Safety concerns with consumer-facing mobile health applications and their consequences: a scoping review. J Am Med Inform Assoc 2020;27(2):330-340 [FREE Full text] [doi: 10.1093/jamia/ocz175] [Medline: 31599936]
79. General data protection regulation (GDPR). Intersoft Consulting. URL: https://gdpr-info.eu/ [accessed 2022-03-17]
80. Mrklas KJ, Barber T, Campbell-Scherer D, Green LA, Li LC, Marlett N, et al. Co-design in the development of a mobile health app for the management of knee osteoarthritis by patients and physicians: qualitative study. JMIR Mhealth Uhealth 2020;8(7):e17893 [FREE Full text] [doi: 10.2196/17893] [Medline: 32673245]
81. Khan ZF, Alotaibi SR. Applications of artificial intelligence and big data analytics in m-health: a healthcare system perspective. J Healthc Eng 2020;2020:8894694 [FREE Full text] [doi: 10.1155/2020/8894694] [Medline: 32952992]
82. Aguilera A, Figueroa CA, Hernandez-Ramos R, Sarkar U, Cemballi A, Gomez-Pathak L, et al. mHealth app using machine learning to increase physical activity in diabetes and depression: clinical trial protocol for the DIAMANTE study. BMJ Open 2020;10(8):e034723 [FREE Full text] [doi: 10.1136/bmjopen-2019-034723] [Medline: 32819981]
83. Burton A. Smartphones versus Parkinson's disease: i-PROGNOSIS. Lancet Neurol 2020;19(5):385-386. [doi: 10.1016/S1474-4422(20)30115-0] [Medline: 32333892]
84. Turchioe MR, Myers A, Isaac S, Baik D, Grossman LV, Ancker JS, et al. A systematic review of patient-facing visualizations of personal health data. Appl Clin Inform 2019;10(4):751-770 [FREE Full text] [doi: 10.1055/s-0039-1697592] [Medline: 31597182]
85. Teo CH, Ng CJ, White A. What do men want from a health screening mobile app? A qualitative study. PLoS One 2017;12(1):e0169435 [FREE Full text] [doi: 10.1371/journal.pone.0169435] [Medline: 28060953]
86. Arora S, Yttri J, Nilse W. Privacy and security in mobile health (mHealth) research. Alcohol Res 2014;36(1):143-151 [FREE Full text] [Medline: 26259009]
87. Clement I, Lorenz A, Ulm B, Plidschun A, Huber S. Implementing systematically collected user feedback to increase user retention in a mobile app for self-management of low back pain: retrospective cohort study. JMIR Mhealth Uhealth 2018;6(6):e10422 [FREE Full text] [doi: 10.2196/10422] [Medline: 29875088]
88. Lin YH, Chen SY, Lin PH, Tai AS, Pan YC, Hsieh CE, et al. Assessing user retention of a mobile app: survival analysis. JMIR Mhealth Uhealth 2020;8(11):e16309 [FREE Full text] [doi: 10.2196/16309] [Medline: 33242023]
89. Adu MD, Malabu UH, Malau-Aduli AE, Drovandi A, Malau-Aduli BS. User retention and engagement with a mobile app intervention to support self-management in Australians with type 1 or type 2 diabetes (My Care Hub): mixed methods study. JMIR Mhealth Uhealth 2020;8(6):e17802 [FREE Full text] [doi: 10.2196/17802] [Medline: 32525491]
90. Wyatt JC. How can clinicians, specialty societies and others evaluate and improve the quality of apps for patient use? BMC Med 2018;16(1):225 [FREE Full text] [doi: 10.1186/s12916-018-1211-7] [Medline: 30501638]
91. Solomon DH, Rudin RS. Digital health technologies: opportunities and challenges in rheumatology. Nat Rev Rheumatol 2020;16(9):525-535. [doi: 10.1038/s41584-020-0461-x] [Medline: 32709998]
92. Santo K, Redfern J. The potential of mHealth applications in improving resistant hypertension self-assessment, treatment and control. Curr Hypertens Rep 2019;21(10):81. [doi: 10.1007/s11906-019-0986-z] [Medline: 31598792]

Abbreviations

AI: artificial intelligence
BMD: bone mineral density
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
ROB: risk of bias

Edited by L Buis; submitted 11.08.21; peer-reviewed by I Tchetverikov, S Van Stee, YL Leung, S Bhattacharjee; comments to author 04.11.21; revised version received 18.12.21; accepted 02.02.22; published 21.04.22.

Please cite as:
Alhussein G, Hadjileontiadis L
Digital Health Technologies for Long-term Self-management of Osteoporosis: Systematic Review and Meta-analysis
JMIR Mhealth Uhealth 2022;10(4):e32557
URL: https://mhealth.jmir.org/2022/4/e32557
doi:10.2196/32557
PMID:35451968

Original Paper

Usability of Smart Home Thermostat to Evaluate the Impact of Weekdays and Seasons on Sleep Patterns and Indoor Stay: Observational Study

Niloofar Jalali[1], MSc, PhD; Kirti Sundar Sahu[1], MPH, MPT; Arlene Oetomo[1], BSc; Plinio Pelegrini Morita[1,2,3,4], MSc, PEng, PhD

[1]School of Public Health and Health Systems, University of Waterloo, Waterloo, ON, Canada

[2]Institute of Health Policy, Management, and Evaluation, University of Toronto, Toronto, ON, Canada

[3]Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada

[4]eHealth Innovation, Techna Institute, University Health Network, Toronto, ON, Canada

Corresponding Author:
Plinio Pelegrini Morita, MSc, PEng, PhD
School of Public Health and Health Systems
University of Waterloo
200 University Avenue West
Waterloo, ON, N2L 3G1
Canada
Phone: 1 519 888 4567 ext 31372
Fax: 1 519 746 6776
Email: plinio.morita@uwaterloo.ca

Abstract

Background: Sleep behavior and time spent at home are important determinants of human health. Research on sleep patterns has traditionally relied on self-reported data. Not only does this methodology suffer from bias but the population-level data collection is also time-consuming. Advances in smart home technology and the Internet of Things have the potential to overcome these challenges in behavioral monitoring.

Objective: The objective of this study is to demonstrate the use of smart home thermostat data to evaluate household sleep patterns and the time spent at home and how these behaviors are influenced by different weekdays and seasonal variations.

Methods: From the 2018 ecobee *Donate your Data* data set, 481 North American households were selected based on having at least 300 days of data available, equipped with ≥6 sensors, and having a maximum of 4 occupants. Daily sleep cycles were identified based on sensor activation and used to quantify sleep time, wake-up time, sleep duration, and time spent at home. Each household's record was divided into different subsets based on seasonal, weekday, and seasonal weekday scales.

Results: Our results demonstrate that sleep parameters (sleep time, wake-up time, and sleep duration) were significantly influenced by the weekdays. The sleep time on Fridays and Saturdays is greater than that on Mondays, Wednesdays, and Thursdays (n=450; *P*<.001; odds ratio [OR] 1.8, 95% CI 1.5-3). There is significant sleep duration difference between Fridays and Saturdays and the rest of the week (n=450; *P*<.001; OR 1.8, 95% CI 1.4-2). Consequently, the wake-up time is significantly changing between weekends and weekdays (n=450; *P*<.001; OR 5.6, 95% CI 4.3-6.3). The results also indicate that households spent more time at home on Sundays than on the other weekdays (n=445; *P*<.001; OR 2.06, 95% CI 1.64-2.5). Although no significant association is found between sleep parameters and seasonal variation, the time spent at home in the winter is significantly greater than that in summer (n=455; *P*<.001; OR 1.6, 95% CI 1.3-2.3). These results are in accordance with existing literature.

Conclusions: This is the first study to use smart home thermostat data to monitor sleep parameters and time spent at home and their dependence on weekday, seasonal, and seasonal weekday variations at the population level. These results provide evidence of the potential of using Internet of Things data to help public health officials understand variations in sleep indicators caused by global events (eg, pandemics and climate change).

(JMIR Mhealth Uhealth 2022;10(4):e28811) doi:10.2196/28811

KEYWORDS

public health; Internet of Things (IoT); big data; sleep monitoring; health monitoring; mobile phone

Introduction

Background

Sleep is vital for human health, as it promotes physical and mental well-being at the individual and population levels [1]. Sleep affects brain function and the performance of other systems within the body such as digestive, cardiovascular, and endocrine [2,3]. The lack of proper sleep can cause fatigue, reduced concentration, and depression [4]. In addition, inefficient or disturbed sleep, due to behaviors such as technology use (eg, use of mobile device screens), can also lead to chronic stress and poor mental health [5]. Researchers have found that insufficient sleep is associated with a significant increase in mortality, diabetes, cardiovascular disease, coronary heart disease, and obesity [6]. Today, sleep health is understood not as an isolated portion of the day but as a significant part of a healthy 24-hour cycle [7].

The Public Health Agency of Canada recommends that individuals aged 18 to 64 years get between 7 and 9 hours of sleep per night and those aged ≥65 years between 7 and 8 hours per night [8]. In Canada, at least one in four adults is not getting enough sleep [8]. Even higher levels of sleep deprivation are reported for those aged 35 to 64 years. Similarly, in the United States, 25% of adults self-reported having <7 to 9 hours of sleep [9]. Insufficient sleep duration occurs despite the time spent indoors having increased over the last century [10].

Sleep research has historically relied on single-sleep questionnaires or sleep diaries [11]. However, self-reported sleep data are prone to recall bias and social desirability bias [11]. Sleep patterns have also been inferred by measuring human brain activity, breathing and blood oxygenation levels, muscle and eye movements, and heart rate [12]. These types of studies, although informative, rely on sleep data collected from artificial laboratory environments and do not reflect sleep patterns in a real-world setting, as individuals are often sleeping in a controlled environment with different, albeit fewer, disturbances. There is a need for the modernization of research methodologies to enable the unbiased collection of sleep pattern data in real-world settings.

The development of smartphones and wearable devices has enabled continuous behavioral monitoring [9]. The assessment of sleep behavior can now be performed using wearable devices such as smartwatches or actigraphs [11] or by interpreting the interactions of a user with the device (eg, smartphone screen) [4].

Advances in smart home technology and the Internet of Things (IoT) have the potential to take behavioral monitoring even further [13]. Smart devices collect data objectively; they are unobtrusive and require zero effort from study participants [14]. This technology has the additional advantage that it can assess behavior in individuals with physical or mental impairments who may be unable to interact with smartphones or wearable devices [15]. These advances offer a previously unprecedented opportunity to monitor sleep behaviors in a real-world setting using methods that can reduce participant bias [16]. Previous studies have successfully used smart thermostats to monitor indoor behavior including sleep [17].

Objectives

The objective of this study is to evaluate the potential use of smart home thermostats to help us understand population-wide sleep patterns, as well as the time the population spends indoors during the year. To assess the impact of different seasonal patterns (eg, days of the week, weekdays vs weekends, and seasons of the year) on sleep health as well as indoor activity, we developed the IoT-based population-level indicators for sleep duration, sleep time, wake-up time, and time spent at home. These indicators are compared across multiple seasonal patterns as weekdays, seasons of the years, and a combined cross-analysis between weekdays and seasons of the year. Ultimately, this study will provide evidence of the potential use of large-scale IoT data to help public health researchers understand the sleep patterns of their population by using the nonobtrusive data collection methods, which will lead to the future use of these data to understand the effects of large-scale global health events (eg, pandemics and climate change).

Methods

This is an exploratory study using secondary data from IoT devices. In this study, we used smart thermostat data from North America.

Data

In this study, our team explored the use of the *Donate Your Data (DYD)* data from the ecobee smart home thermostat. The data are composed of the anonymized indoor activity of households captured every 5 minutes through the embedded motion sensors [18]. Approximately 98% of participating households in the DYD program are in North America. Taking the specification of ecobee motion sensors into account, for a family size of up to 4, the distribution of floor area has been identified. On the basis of previous exploratory work, we identified the optimum number of sensors based on the average floor area and minimum distance coverage of the sensors. Our household selection criteria included a minimum of 300 days of data available on the data set, the presence of at least six motion sensors in the home, and a maximum of 4 residents. The data management and analysis have been performed on Microsoft Azure Databricks and the scikit learn library [19] in Python (version 3.6).

Defining the Sleep Parameters and Time Spent at Home

The original DYD data were reported every 5 minutes. To avoid unnecessary uncertainty, the data were aggregated in 30-minute intervals [18,20,21]. The sum of activation of all sensors, in every 30-minute interval, corresponds to the activity level for that period.

The activation of 1 sensor for 5 minutes corresponds to a score of 1. In a 30-minute interval, the active interval was defined as a score ≥4 (eg, 1 sensor active for 4×5-minute interval [20 minutes], 4 sensors each active for 5 minutes, or any combination of the aforementioned parameters). Intervals with activation sums below this threshold were considered noise. The data were compounded into a binary vector representing a daily record with 48 time slots [18].

This activation pattern was identified to ensure that, while avoiding unnecessary noise, two types of behavior can be picked up by the system: (1) individuals staying in the same room for extended periods, hence activating one sensor sequentially, or (2) individuals moving around the house, hence activating multiple sensors in a shorter time frame.

To develop the different sleep indicators discussed above, we divided each day into two parts, namely, (1) midnight until noon and (2) 8 PM to midnight, and disregarded the time interval from noon to 8 PM. In every 2 consecutive days (eg, days 0 and 1), a sleep cycle was defined as the second part of day 0 combined with the first part of day 1 (Figure 1).

To identify the sleep indicators of each household, the following steps were performed:

1. using the Gaussian mixture model [22] to segment the sleep cycle records into different clusters to differentiate the sleep–wake-up behaviors through the selected time scale,
2. identifying the sleep–wake-up patterns in each cluster by averaging the activation of sensors at each time interval (if the average of activation is >0.5, it is assumed that the sensor was active at that time; otherwise, it is assumed as inactive), and
3. specifying the sleep indicators for each sleep–wake-up pattern, using the following assumptions:
    - the deactivation (sleep time) occurs before activation (wake-up time),
    - the earliest possible deactivation (sleep time) can start at 8 PM, and
    - the largest interval between 2 consecutive deactivation and activation times (eg, starts at 8 PM until noon) represents the sleep time, wake-up time, and sleep duration.

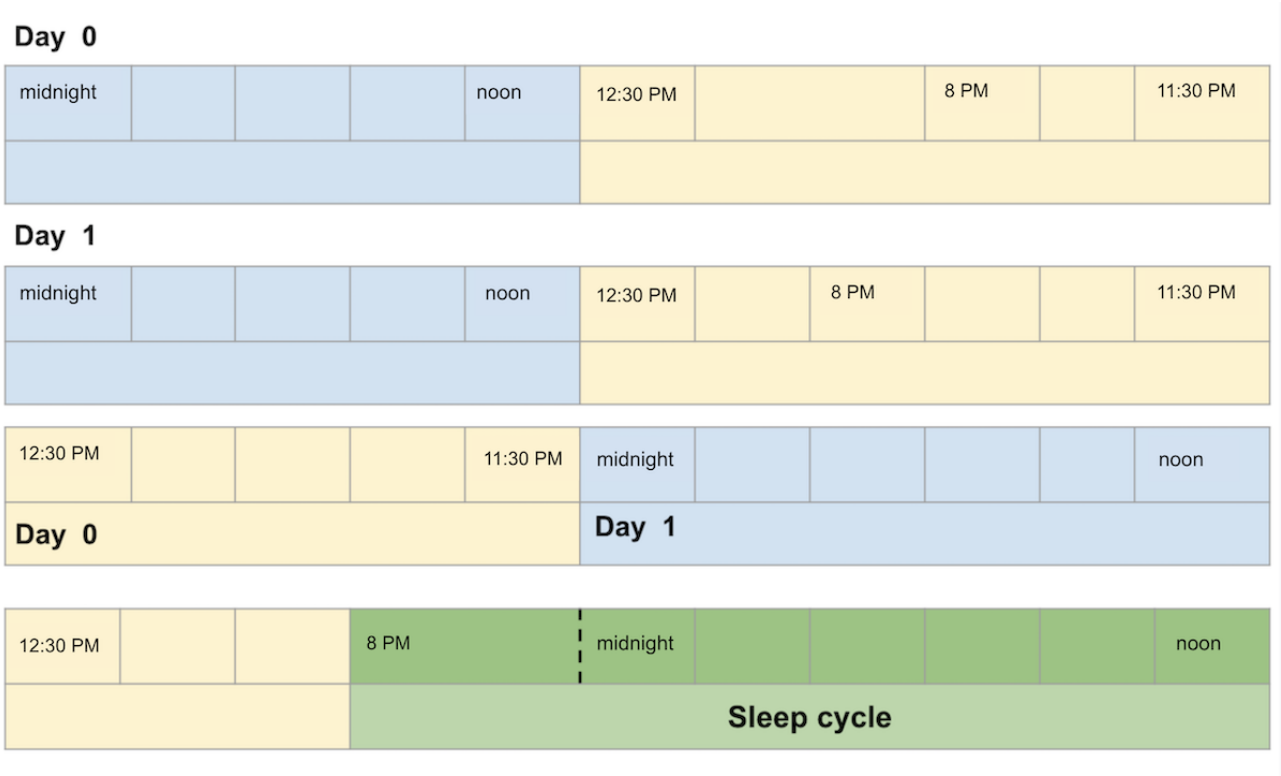
Ultimately, the weighted average of each indicator demonstrates the result of sleep parameters for each household at the selected time scale, where the weighted average is defined by the following:

In addition to sleep indicators, we also explored the amount of time spent at home, where the daily cycle records are used to identify the different patterns in each cluster. From each pattern, the amount of time spent at home is defined by the sum of activation of sensors at each interval. The weighted average of each cluster demonstrates the ultimate result.

To explore some of the seasonal effects on sleep and indoor activity patterns, we stratified the data based on different time frames: weekdays and seasons of the year. Next, we compared the different indicators in each of the different time frames (ie, sleep and time spent at home). For each stratified group, the indicators were calculated and the statistical significance between groups was evaluated using analysis of variance (ANOVA) [23]. The statistically significant differences of two indicators (ie, sleep and time at home) were further explored using Tukey post hoc tests [24].

For each of the stratified groups, we present complete descriptive statistics: sample size, mean (SD), SE, and 99% CI of the mean. We assumed that the subsets are independent and distributed normally and the variances are homogeneous [25]. The homogeneity of variance was evaluated using the Levene test [25].

Figure 1. Combining the 2 consecutive dates to create the sleep cycle and identify the sleep time, sleep duration, and wake-up time.



Seasons of the Year

Data were stratified by seasons of the year, dividing the annual record of each household into 4 seasons [26]. The start and end dates of each season are as follows: winter (January 1 to March 21), spring (March 22 to June 21), summer (June 22 to September 21), and fall (September 22 to December 21).

The distribution of all the subsets as well as the homogeneity of variances was checked using the Levene test [23,27]. The 1-way ANOVA and Tukey post hoc tests were used to evaluate the statistically significant relationship between sleep time, wake-up time, sleep duration, and time spent at home with respect to different seasons.

Weekdays

A similar approach was used for stratifying the data on different weekdays. The annual records of each household were stratified into 7 subsets, representing data from each weekday. The distribution of all the subsets, as well as the homogeneity of variances, was investigated using the Levene test. The 1-way ANOVA and Tukey post hoc tests were used to evaluate the statistically significant relationship among sleep time, wake-up time, sleep duration, and time spent at home when comparing different weekdays.

Seasons of the Year and Weekdays

In the combined analysis exploring seasons of the year and weekdays, the annual record of each household was divided into two independent variables: each of 4 seasons and each of 7 days of the week.

The same approach has been replicated for the analysis of different sleep parameters and time spent at home. The 2-way ANOVA [23] and Tukey post hoc test were used to compare the differences of each parameter for different seasons and weekdays simultaneously.

Results

Overview

To examine sleep patterns across large populations, we explored the *DYD* ecobee smart home thermostat data set. The DYD program is hosted by ecobee and provides researchers with access to anonymized data from 110,000 households. After we assessed these households for eligibility, a total of 481 households met the inclusion criteria and were included in this study: they had at least 300 days of data available in the DYD data set, at least six sensors, and a maximum of 4 residents. Of the 481 households included in the study, 390 (81.1%) were in the United States, 63 (13.1%) were in Canada, and 28 (5.8%) had undeclared locations. The largest proportion (40/390, 10%) of households in the United States were in the state of California. In Canada, most households were from the province of Ontario (40/63, 65%).

Effect of Seasons of the Year on Sleep Parameters and Time Spent at Home

To examine the effect of different seasons on sleep parameters, household data were stratified into 4 seasons. For each of the four indicators (sleep time, wake-up time, sleep duration, and time spent at home), the statistical distributions are presented in Table 1.

The sleep duration and time spent at home, is presented as the total number of minutes. While, the sleep time and wake-up time are presented using the standard hh:mm format.

Knowing that the null hypothesis of the Levene test is that the groups we are comparing all have equal population variances, the results would confirm the homogeneity of variances in each of the stratified groups is (Table 2).

We can declare all groups homogeneous with a significance threshold of (P<.01). A 1-way ANOVA test was used to compare the seasonal differences for different sleep parameters and time spent at home (Table 3).

Assuming that all other variables are constant, there is a statistically significant difference in time spent at home in different seasons of the year. However, season alone has no statistically significant impact on sleep time, wake-up time, and sleep duration.

The Tukey post hoc test was performed to identify the significant pairs. The difference in means, CIs, and adjusted P values per pair are presented in Table 4.

The results indicate a significant difference in the time spent at home between winter and summer (P<.001; odds ratio [OR] 1.6, 95% CI 1.3-2.3).

The results of the Tukey post hoc test demonstrate that the time spent at home among these households during the winter is statistically significant from that during the summer. On average, individuals in these households spend, in the summer, 1 extra hour outside when compared with that in the winter. These results demonstrate the potential of this IoT data set to inform public health practice by providing insights on population-level behaviors in different conditions.

Table 1. Descriptive statistics for season-based stratified groups for sleep parameters (sleep time, wake-up time, and sleep duration) and time spent at home.[a]

Season	Value, N	Value, mean (SD; SE; 99% CI)
Sleep time		
Spring	461	22:32 (98.5; 4.59; 22:20-22:44)
Summer	455	22:32 (108.37; 5.08; 22:19-22:45)
Fall	461	22:25 (89.12; 4.15; 22:14-22:36)
Winter	466	22:35 (107.11; 4.96; 22:22-22:48)
Wake-up time		
Spring	461	6:22 (93.05; 4.33; 6:11-6:33)
Summer	455	6:31 (84.75; 3.97; 6:21-6:41)
Fall	461	6:26 (88.36; 4.12; 6:15-6:37)
Winter	466	6:26 (102.75; 4.76; 6:14-6:39)
Sleep duration		
Spring	461	466.54 (106.23; 4.95; 453.79-479.28)
Summer	455	472.54 (109.59; 5.14; 459.31-485.77)
Fall	461	477.83 (106.2; 4.95; 465.09-490.57)
Winter	466	465.05 (110.65; 5.13; 451.84-478.25)
Time spent home		
Spring	461	523.13 (227.56; 10.6; 495.83-550.43)
Summer	455	486.58 (237.4; 11.13; 457.91-515.25)
Fall	461	529.39 (226.42; 10.55; 502.22-556.55)
Winter	466	546.43 (235.53; 10.91; 518.32-574.53)

[a]The mean of sleep time and wake-up time, along with their CIs, is presented using a standard hh:mm format. The SD of all 4 indicators, as well as the sleep duration and time spent at home, is presented as the total number of minutes.

Table 2. Evaluating the homogeneity of variances in different seasonal groups using the Levene test.

Indicator	P value
Wake-up time	.47
Sleep time	.12
Sleep duration	.84
Time spent home	.66

Table 3. Analysis of variance test to explore the effect of seasons of the year on sleep time, sleep duration, wake-up time, and time spent at home.

	Sum of squares	*df*	*F* test	*P* value
Sleep time				
C(season)	26,092.90	3	0.85	.47
Residual	18,783,017.85	1839		
Wake-up time				
C(season)	20,683.21	3	0.81	.49
Residual	15,744,147.50	1839		
Sleep duration				
C(season)	13.26	3	1.36	.25
Residual	5979.04	1839		
Time spent home				
C(season)	874,376.67	3	5.43	<.001
Residual	98,783,401.15	1839		

Table 4. The Tukey post hoc test comparing the statistical significance of the difference between each pair of seasons.

	Group 1	Group 2	Difference (SE)	95% CI of the mean	*q* value	*P* value
Sleep time						
	Winter	Spring	3.17 (−4.29)	−13.9 to 20.24	0.68	.90
	Winter	Summer	2.89 (2.96)	−14.24 to20.01	0.61	.90
	Winter	Fall	10.21 (−13.54)	−6.86 to 27.28	2.17	.42
	Summer	Spring	0.28 (−7.25)	−16.89 to 17.46	0.06	.90
	Fall	Spring	7.04 (9.25)	−10.08 to 24.15	1.49	.69
	Fall	Summer	7.32 (16.49)	−9.85 to 24.49	1.55	.67
Wake-up time						
	Winter	Spring	4.71 (−10.26)	−10.92 to 20.34	1.10	.85
	Winter	Summer	4.78 (−17.33)	−10.9 to 20.46	1.11	.85
	Winter	Fall	0.47 (−12.22)	−15.16 to 16.10	0.11	.90
	Summer	Spring	9.49 (7.07)	−6.24 to 25.21	2.19	.41
	Fall	Spring	4.24 (1.96)	−11.43 to 19.91	0.98	.89
	Fall	Summer	5.25 (−5.11)	−10.48 to 20.97	1.21	.80
Sleep duration						
	Winter	Spring	1.49 (−4.42)	−16.79 to 19.76	0.30	.90
	Winter	Summer	7.49 (−1.06)	−10.84 to 25.83	1.49	.70
	Winter	Fall	12.78 (−4.45)	−5.50 to 31.05	2.54	.27
	Summer	Spring	6.00 (−3.36)	−12.38 to 24.39	1.19	.81
	Fall	Spring	11.29 (0.03)	−7.03 to 29.61	2.24	.39
	Fall	Summer	5.29 (3.39)	−13.1 to 23.67	1.05	.87
Time spent home						
	Winter	Spring	23.3 (−7.65)	−15.85 to 62.45	2.16	.42
	Winter	Summer	59.85 (−1.90)	20.57 to 99.13	5.54	<.001
	Winter	Fall	17.04 (−10.18)	−22.11 to 56.19	1.58	.66
	Summer	Spring	36.55 (−5.75)	−2.83 to 75.94	3.38	.08
	Fall	Spring	6.26 (2.54)	−33.00 to 45.51	0.58	.90
	Fall	Summer	42.81 (8.29)	3.43 to 82.19	3.95	.03

Effect of Different Weekdays Sleep Parameters and Time Spent at Home

The ability to identify sleep indicators and time spent at home on different weekdays, using an IoT data set, is further evidence of the potential use of these data to understand and monitor the behaviors of a population [28]. To examine the influence of different weekdays on the 4 indicators, household data were divided between 7 weekdays.

For each indicator, the descriptive statistics for each weekday as well as the homogeneity of variances within different weekdays are presented in Tables 5 and 6, respectively.

We found no significant difference of variances among the 7 weekdays for wake-up time, sleep time, and time spent at home. However, the sleep duration is not fulfilling the variance homogeneity assumption, and the results need to be generalized with precaution. As explained in Figure 1, a typical sleep cycle is spread across 2 days, beginning and ending on different dates. Therefore, the sleep time and sleep duration occur on one day and wake-up time on the next day (Figure 1), which is likely one of the reasons for the nonhomogeneous distribution of variances. The sleep time, wake-up time, sleep duration, and time spent at home on weekdays are illustrated in Figure 2.

The average sleep time for the entire sample was 10:40 PM (Figure 2A). Most households had a sleep time earlier than 10:40 PM on Mondays, Tuesdays, Wednesdays, Thursdays, and Sundays. However, 50.1% (232/463) and 52.8% (245/464) of the households had sleep time of equal to or later than 10:40 PM on Fridays and Saturdays, respectively.

The average wake-up time of all the households in the entire sample was 6:20 AM (Figure 2B). Most households have a wake-up time earlier than 6:20 AM during weekdays, and the average wake-up time on weekends was greater than 6:20 AM (Figure 2B).

The average sleep duration for the entire sample was 8 hours (Figure 2C). The average sleep duration of the households was 7½ hours on Mondays, Tuesdays, Wednesdays, Thursdays, and Sundays. The sleep duration was longer than 8 hours on Fridays and Saturdays.

The average time spent at home for the entire sample was 9 hours (Figure 2D). The time spent at home on weekends was >9 hours.

The 1-way ANOVA showed statistically significant differences in sleep indicators and time spent at home for the stratified weekdays, assuming that all other variables were constant (Table 7).

The Tukey post hoc test comparing sleep time, sleep duration, wake-up time, and time spent at home for different weekday pairs are presented in Table 8. The results of the overall comparison and the significant pairs are presented in Multimedia Appendix 1 and Table 8, respectively. Owing to a large number of comparisons, we only present the statistically significant results in Table 8. Out of all the 21 possible pair combinations for each indicator, we had 8 (38%) statistically significant pairs when comparing sleep time, 10 (47%) when comparing wake-up time, 10 (47%) when comparing sleep duration, and 6 (29%) when comparing time spent at home.

The Tukey post hoc test provides evidence that the sleep time on Fridays and Saturdays was statistically different from that on Mondays, Wednesdays, and Thursdays (OR 1.8, 95% CI 1.5-3; P<.001). The most significant difference in sleep time was between Mondays and Saturdays, with an average of 40 minutes earlier on Mondays than on Saturdays (Table 8).

There was a statistically significant difference in wake-up time on Saturdays and Sundays compared with that on the remaining weekdays (OR 5.6, 95% CI 4.3-6.3; P<.001). The maximum wake-up time difference was between Tuesdays and Sundays, with households waking an average of 76 minutes later on Sundays (Table 8).

The sleep duration on Fridays and Saturdays was statistically significant from the other days of the week (OR 1.8, 95% CI 1.4-2; P<.001), with households sleeping longer on Fridays and Saturdays than on the other days of the week. The highest sleep duration difference was between Tuesdays and Saturdays, with an extra 36 minutes of sleep on Saturdays (Table 8).

There was a statistically significant difference in the time spent at home on Sundays with respect to other days of the week (OR 2.06, 95% CI 1.64-2.5; P<.001). The highest difference in time spent at home is between Thursdays and Sundays, with an average of 96 minutes more time spent at home on Sundays (Table 8).

These results indicate that the data collected by IoT smart home sensors can provide evidence of expected differences between sleep time, wake-up time, sleep duration, and time spent at home. These results provide further evidence for using these data to monitor population-level changes caused by global events.

Table 5. Descriptive statistics of stratified weekday group for sleep indicators and time spent at home.[a]

	Value, N	Value, mean (SD; SE; 99% CI)
Sleep time (PM)		
Monday	463	10:26 (81.3; 3.78; 10:16-10:35)
Tuesday	459	10:35 (94.53; 4.41; 10:23-10:46)
Wednesday	458	10:31 (86.22; 4.03; 10:20-10:41)
Thursday	460	10:33 (99.67; 4.65; 10:21-10:44)
Friday	463	10:54 (89.2; 4.15; 10:43-11:04)
Saturday	464	11:03 (104.04; 4.84; 10:51-11:16)
Sunday	466	10:36 (85.27; 3.95; 10:26-10:47)
Wake-up time (AM)		
Monday	463	6:06 (70.6; 3.28; 5:57-6:14)
Tuesday	459	5:54 (74.22; 3.46; 5:45-6:02)
Wednesday	458	6:03 (73.7; 3.44; 5:55-6:12)
Thursday	460	6:04 (84.86; 3.96; 5:54-6:14)
Friday	463	6:05 (72.76; 3.38; 5:57-6:14)
Saturday	464	7:00 (101.23; 4.71; 6:48-7:12)
Sunday	466	7:08 (79.51; 3.68; 6:58-7:17)
Sleep duration		
Monday	463	449.1 (95.83; 4.45; 437.63-460.57)
Tuesday	459	447.76 (97.47; 4.55; 436.04-459.48)
Wednesday	458	449.88 (92.13; 4.3; 438.79-460.97)
Thursday	460	449.73 (96.21; 4.49; 438.17-461.28)
Friday	463	477.99 (110.99; 5.16; 464.7-491.27)
Saturday	464	483.25 (111.16; 5.17; 469.92-496.57)
Sunday	466	450.18 (89.15; 4.13; 439.54-460.82)
Time spent home		
Monday	463	516.47 (216.85; 10.08; 490.51-542.43)
Tuesday	459	486.14 (215.28; 10.05; 460.26-512.02)
Wednesday	458	503.46 (219.69; 10.27; 477.02-529.91)
Thursday	460	482.41 (213.16; 9.94; 456.81-508.01)
Friday	463	498.95 (223.06; 10.37; 472.25-525.65)
Saturday	462	533.51 (215.45; 10.02; 507.69-559.33)
Sunday	466	576.27 (220.66; 10.22; 549.94-602.6)

[a]The mean of sleep time and wake-up time, along with their CIs, is presented using a standard hh:mm format. The SD of all 4 indicators, as well as the sleep duration and time spent at home, is presented as the total number of minutes.

Table 6. Evaluating the homogeneity of variances in different weekday groups using the Levene test.

Indicator	*P* value
Wake-up time	.02
Sleep time	.02
Sleep duration	<.001
Time spent home	.86

Figure 2. The box plot of wake-up time (A), sleep time (B), sleep duration (C), and time spent at home (D) for weekdays. The dashed line represents the total average.

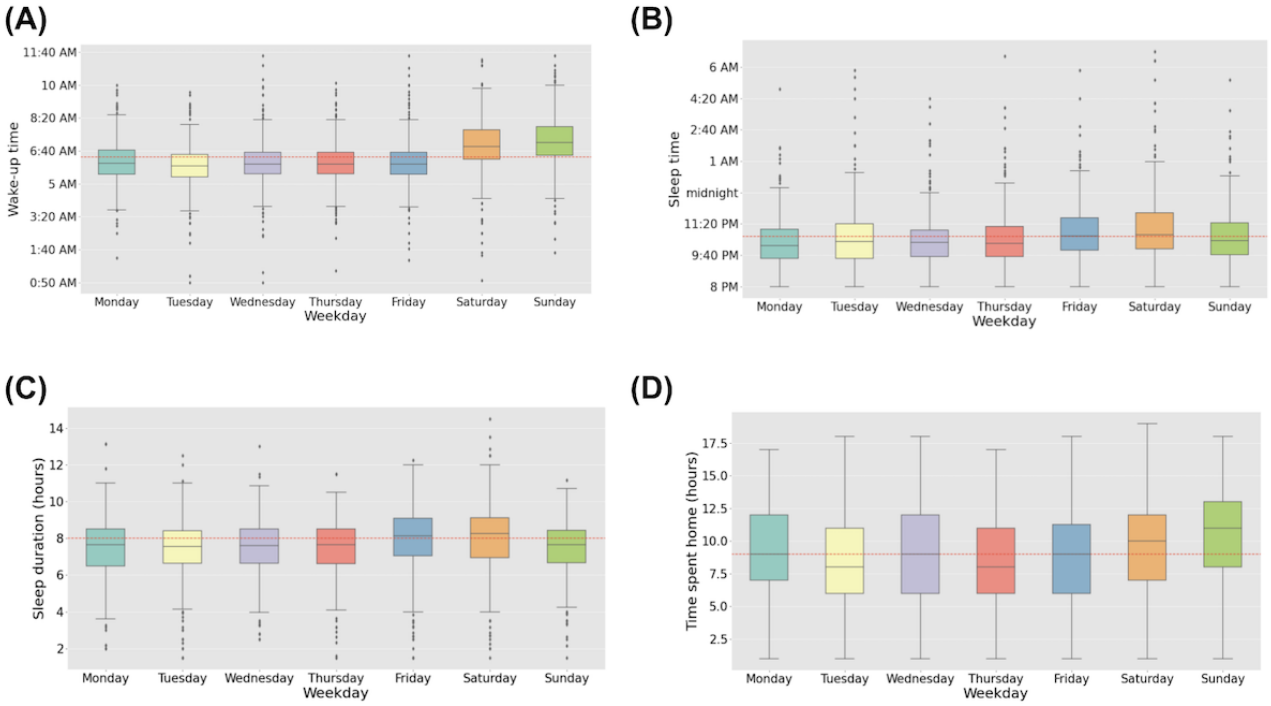


Table 7. The analysis of variance test compares weekdays' impact on sleep time, sleep duration, wake-up time, and time spent at home.

	Sum of squares	*df*	*F* test	*P* value
Sleep time				
C(wday)[a]	524,447.23	6	10.39	<.001
Residual	27,150,245.54	3226		
Wake-up time				
C(wday)	2,591,512.52	6	67.05	<.001
Residual	20,782,158.97	3226		
Sleep duration				
C(wday)	180.43	6	10.97	<.001
Residual	8843.10	3226		
Time spent home				
C(wday)	2,970,000	6	10.39	<.001
Residual	15,300,000,000	3224		

[a]wday: weekday.

Table 8. The significant results of the Tukey test for sleep time, wake-up time, sleep duration, and time spent home were stratified by different weekdays.

Group 1	Group 2	Difference (SE)	95% CI of the mean	*q* value	*P* value
Sleep time					
Monday	Friday	29.69 (9.93)	10.34-45.93	6.60	.001
Monday	Saturday	40.35 (25.11)	20.13-55.74	8.89	.001
Tuesday	Saturday	28.45 (9.15)	11.10-46.78	6.77	.001
Wednesday	Friday	23.23 (2.74)	5.15-40.83	5.38	.003
Wednesday	Saturday	33.90 (17.91)	14.94-50.64	7.66	.001
Thursday	Friday	21.93 (−9.80)	3.34-38.98	4.95	.007
Thursday	Saturday	32.60 (5.38)	13.13-48.80	7.24	.001
Saturday	Sunday	−25.67 (−17.14)	9.36-44.91	6.37	.001
Wake-up time					
Monday	Saturday	51.59 (23.42)	38.89-70.00	14.61	.001
Monday	Sunday	61.23 (8.25)	46.73-77.77	16.73	.001
Tuesday	Saturday	66.01 (20.75)	50.9-82.08	17.80	.001
Tuesday	Sunday	75.65 (5.59)	58.74-89.84	19.93	.001
Wednesday	Saturday	55.36 (20.83)	40.98-72.17	15.13	.001
Wednesday	Sunday	65.00 (5.66)	48.81-79.93	17.26	.001
Thursday	Saturday	54.63 (8.48)	40.38-71.53	14.99	.001
Thursday	Sunday	64.27 (−6.69)	48.21-79.3	17.11	.001
Friday	Saturday	53.40 (20.32)	39.13-70.23	14.67	.001
Friday	Sunday	63.04 (5.15)	46.96-78.00	16.80	.001
Sleep duration					
Monday	Friday	28.32 (16.06)	9.63-48.15	6.26	.001
Monday	Saturday	32.07 (17.06)	14.88-53.42	7.39	.001
Tuesday	Friday	32.67 (14.72)	10.93-49.53	6.54	.001
Tuesday	Saturday	36.42 (15.73)	16.18-54.8	7.67	.001
Wednesday	Friday	29.19 (19.87)	8.79-47.42	6.07	.001
Wednesday	Saturday	32.94 (20.88)	14.04-52.68	7.20	.001
Thursday	Friday	28.42 (16.14)	8.97-47.55	6.11	.001
Thursday	Saturday	32.17 (17.14)	14.22-52.82	7.25	.001
Friday	Sunday	−29.79 (−21.29)	8.58-47.04	6.03	.001
Saturday	Sunday	−33.54 (−22.29)	13.83-52.30	7.17	.001
Time spent home					
Monday	Sunday	57.96 (1.06)	17.64-101.95	5.92	.001
Tuesday	Sunday	89.32 (1.85)	47.88-132.38	8.90	.001
Wednesday	Sunday	73.31 (−2.23)	30.53-115.08	7.19	.001
Thursday	Saturday	53.82 (0.72)	8.79-93.42	5.04	.004
Thursday	Sunday	95.69 (5.39)	51.63-136.08	9.27	.001
Friday	Sunday	79.39 (−5.58)	35.16-119.47	7.65	001

The Effect Combined of Seasons and Weekdays on Sleep Indicators and Time Spent at Home

To further investigate the ability of the smart thermostat IoT data to differentiate patterns in the data and days with unique behavioral patterns, we investigated the combined effect of weekdays and seasons of the year. This proposed analysis will focus on comparing the 4 indicators on different seasons of the year but blocking out analysis by weekday. We first divided the data into 7 weekdays and then, within each weekday, into 4 seasons. Descriptive statistics of the season-weekday groups, as well as the validated results of homogeneity of variances among all subsets for wake-up time, sleep duration, and time spent at home have been provided in Multimedia Appendix 1.

We found no significant difference of variance among subsets, except for the Saturday sleep time, which requires caution when generalizing the results.

The 2-way ANOVA [29] demonstrates the statistically significant differences in sleep indicators and time spent at home with respect to the variation of seasons and weekdays (Table 9).

The greatest impact of seasonality on weekday-specific variations was seen with respect to time spent at home. Figure 3 illustrates the differences through box plots for Thursdays (Figure 3A), Fridays (Figure 3B), Saturdays (Figure 3C), and Sundays (Figure 3D) with respect to the different seasons. Through ANOVA, followed by the Tukey post hoc test, there was a statistically significant difference in time spent at home on Thursdays in the summer in contrast with other seasons. The average time households spend at home on Thursdays is 8½ hours. The time spent at home during the summer was significantly less than that in all other seasons (Figure 3A–D). The time spent at home in the summer on Fridays (Figure 3B), Saturdays (Figure 3C), and Sundays (Figure 3D) is significantly less than in the winter. The time spent at home on Saturdays in the summer is statistically different from that in winter and fall. Households spend less time at home on Saturdays during the summer than during the fall (Figure 3C).

The results demonstrate that the sleep indicators and the time spent at home are significantly associated with variation of seasons of the year and weekdays. To identify the season's impact on each weekday and compare the different parameters of sleep and time spent at home, we used the Tukey post hoc test to compare the variation of seasons with respect to the specific weekday. The overall comparison and the significant pair results are presented in the Multimedia Appendix 1 and Table 10, respectively.

Different seasons did not have a statistically significant effect on sleep time and sleep duration. There was a statistically significant difference between the wake-up time on Fridays between summer and winter, with an average of 27 (SD 5) minutes (Table 10).

Table 9. The 2-way analysis of variance test to compare season and weekdays' impact on sleep time, sleep duration, wake-up time, and time spent at home.

	Sum of squares	*df*	*F* test	*P* value
Sleep time				
C(wday)[a]	1,639,363	6	32.18	<.001
C(season)	136,515.4	3	5.36	.001
Residual	112,487,100	13,247.0		
Wake-up time				
C(wday)	5,786,750.05	6	134.32	<.001
C(season)	379,805.86	3	17.63	<.001
Residual	95,119,438.37	13,247.0		
Sleep duration				
C(wday)	394.48	6	22.47	<.001
C(season)	135.76	3	15.47	<.001
Residual	38.752.79	13,247.0		
Time spent home				
C(wday)	9,840,000	6	38.08	<.001
C(season)	4,550,000	3	35.19	<.001
Residual	571,000,000	13,247.0		

[a]wday: weekday.

Figure 3. The box plot of time spent at home with respect to different seasons: Thursday seasons (A); Friday seasons (B); Saturday seasons (C); and Sunday seasons (D). The dashed line represents the total average.

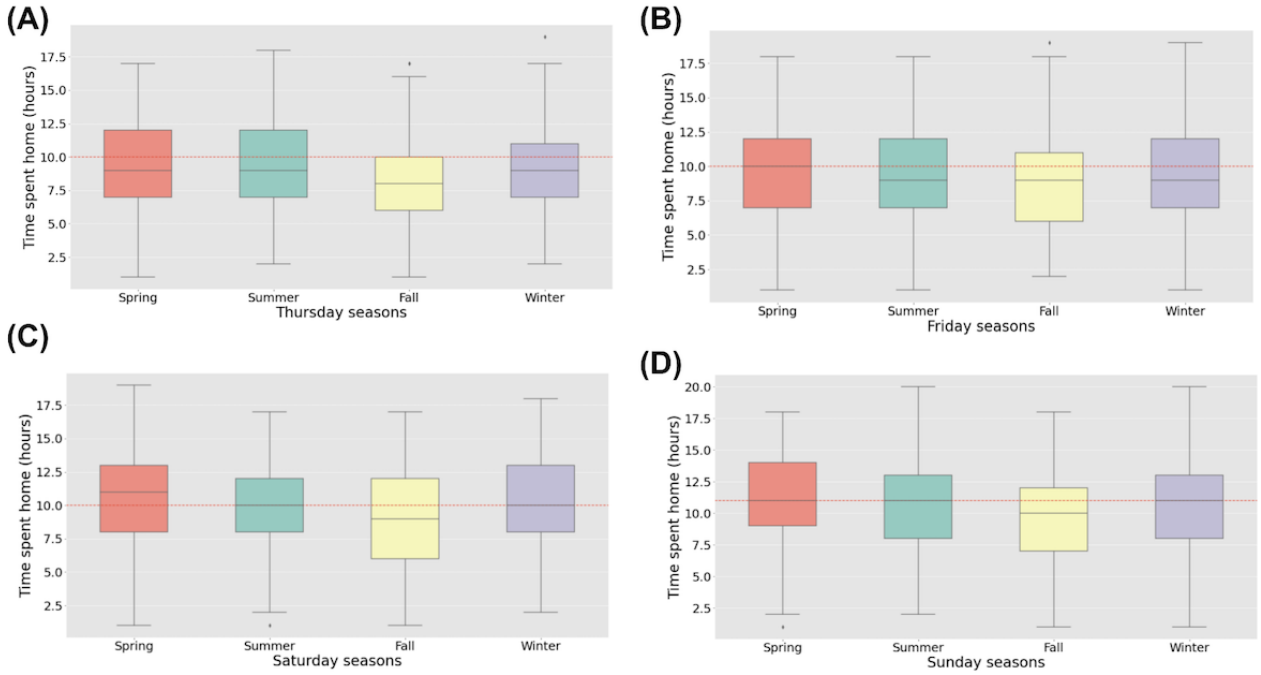


Table 10. The significant results of the Tukey test for sleep time, wake-up time, sleep duration, and time spent at home for each weekday but different seasons.

Group 1	Group 2	Difference (SE)	95% CI of the mean	*q* value	*P* value
Wake-up time					
(Winter, Friday)	(Summer, Friday)	27.06 (4.93)	3.48-44.42	6.15	<.001
Time spent home					
(Winter, Thursday)	(Summer, Thursday)	−81.23 (−5.77)	29.61-130.07	8.35	<.001
(Winter, Friday)	(Summer, Friday)	−60.80 (−3.02)	8.06-108.36	6.10	.01
(Winter, Saturday)	(Summer, Saturday)	−81.37 (−7.01)	32.01-132.58	8.60	<.001
(Winter, Sunday)	(Summer, Sunday)	−69.50 (−5.42)	21.78-122.24	7.53	<.001
(Spring, Thursday)	(Summer, Thursday)	−74.08 (−4.07)	24.00-124.36	7.77	<.001
(Summer, Thursday)	(Fall, Thursday)	58.16 (−0.72)	10.91-111.00	6.40	<.001
(Summer, Saturday)	(Fall, Saturday)	63.58 (−11.09)	11.55-111.90	6.46	<.001

Discussion

Principal Findings

Overview

The advent of smart home technology has provided a previously unprecedented opportunity to collect population-wide, reliable, objective, nonintrusive data on human behavioral patterns. In this study, we evaluated the potential use of ecobee smart home thermostat data as a potential data source for informing public health practice.

Validation of IoT Smart Home Data as Indicators of Healthy Behavior

The initial step in the validation of a new data source as an indicator for public health monitoring is to demonstrate the discriminability of the data in the data set [30]. As new indicators are developed, researchers must demonstrate that variations in independent variables (ie, days of the week and seasons) will result in consistent and expected changes in the dependent variables (ie, sleep indicators and time at home) [31,32]. In this study, we successfully demonstrated the variations in our public health indicators of sleep and time at home (Table 8) caused by changes in the dependent variables. In the next few sections, we provide further discussion of the potential benefits of these indicators for public health practice and research.

Sleep Parameters and Their Impact on Public Health

Our data showed that the seasons of the year had no impact on sleep time, wake-up time, and sleep duration, assuming that all other variables are constant. Our results are in agreement with previous sleep studies conducted in contemporary Western societies [8,33-36] and indicate that the use of smart home thermostat data is a valid method to examine household sleep patterns. In contrast, a 2018 study conducted in Japan showed that seasons significantly affect the sleep parameters of the adult population [37]. These differences may reflect geo-climatic and sociocultural differences and their potential effects on sleep parameters.

Consistent with other reports [38], we found a strong influence of weekdays on sleep time, wake-up time, and sleep duration (Table 8). The sleep time and sleep duration on Fridays and Saturdays were significantly greater than on the other days of the week (Table 8), which is consistent with other studies in this space [38,39]. There is a statistically significant difference in wake-up time between weekends and weekdays, as also demonstrated by Zhang et al [40]. Understanding influences on sleep patterns is an important health determinant, as short sleep duration on weekdays (weekday sleep debt) is a risk factor for chronic diseases and can lead to early mortality [38].

Time Spent Indoors and Its Impact on Public Health

Using the ecobee thermostat data, we also demonstrated that residents spend on average 9 hours per day in their dwellings. Time spent at home varied significantly with respect to different days of the week and season. Previous studies on the impact of different seasons and time spent at home identify that time at home outdoors can vary according to factors such as the seasons, occupation, and age [41,42].

Our results demonstrate a statistically significant difference in the time spent at home between summer and winter, assuming that all other variables are constant (Table 4). In winter, the average daily time spent at home is 1 hour greater than that during the summer (Table 1). Previous studies are consistent with our results, showing a significant difference between physical activity and sedentary behavior between the winter and summer seasons, increasing time indoors during the winter, also found in the literature [33,43-45].

The time spent indoors varies from <8 hours to >10 hours for different days of the week (from a minimum average of 482.41 minutes on Thursdays to a maximum average of 576.27 minutes on Sundays) (Table 5), signifying a strong influence of weekdays and weekend variation on indoor behavior. Significantly more time was spent at home on Sundays than on other days of the week. These effects are consistent with the literature, as demonstrated by Bittman [46].

The duration of staying at home for each individual is linked to physical and mental health [47-49]. More time spent inside the house causes detachment from the natural world, reduced sunlight exposure, lower probability of physical activity, sedentary behavior, exposure to air pollutants, and reduced social interaction [50-52]. Physical activity and sedentary behavior have a strong influence on the risk factors of chronic diseases [53].

Seasonal Weekdays Scale

This study is the first to explore the simultaneous effect of seasons and weekdays on sleep parameters and time spent indoors. Our findings show a significant effect on wake-up time and time spent at home, but not on sleep time and sleep duration (Table 10). We identified that the wake-up time on Fridays in the summer is significantly greater than in the winter (Table 10). The average time spent at home in the summer (Thursdays, Fridays, Saturdays, and Sundays) is significantly less than that in the winter, which is consistent with findings from other researchers such as Plasqui and Westerterp [54], Matz et al [55], and Farrow et al. [56]. Residents spend less time at home on Saturdays during the summer than during the fall.

Strengths of IoT-Based Public Health Monitoring

Previous studies used self-reported surveys [57], sleep diaries [58], accelerometers [59], wearables [60], and social media [61] to calculate different sleep parameters at the individual and population levels. This study is unique in its use of smart thermostats and remote motion sensors as data sources. The sleep results presented here are consistent with those published using other data sources, highlighting the validity of our methods. These data sets and analysis techniques can monitor behavioral health risk indicators at the population level for different geographic and geopolitical locations.

The use of big data from IoT devices has several advantages over traditional data sources [62]. This analysis used 1 year of data that can be extended to include additional years for the same households, enabling identification of longitudinal, long-term patterns. The total data set includes data from >100,000 households, with >5 years of data. Another strength of this data set is the granularity of the data, which are reported every 5 minutes. The use of zero-effort technology and data integration from several sensors will further enhance public health syndromic surveillance [63], with near–real-time monitoring potentially providing evidence of the immediate effect of public health policies at the population level. The COVID-19 pandemic has provided strong evidence of how our monitoring systems in public health need to be optimized and that the gap between data collection and data use must be reduced [64].

Conclusions

This study is the first of its kind, leveraging smart home thermostat data to monitor sleep indicators and time spent at home on different temporal scales (weekdays, seasonal, and seasonal weekdays) at the population level. This approach not only uses nonadhesive and zero-effort technology to collect the data but can also improve the decision-making of public health officials through large-scale, near–real-time IoT data. Our results demonstrate the variations in sleep indicators and time spent at home for different days of the week and different seasons of the year, which provides evidence of the discriminability of the data and will potentially lead to wider-scale use of IoT data in public health, which has the potential to monitor the effects of global events such as climate change and pandemics.

Limitations

Some of the limitations of this study include the absence of household socioeconomic and demographic data. The DYD data set is collected and compiled by ecobee, which unfortunately prevents us from collecting important additional data from our participants. As a result, no stratified analysis is possible for age, sex, race, and health indicators. In addition, the self-reported number of occupants and the size of the house may be inaccurate. As this study uses data at the household level, an individual-level analysis is not possible. Similarly, the data provided by ecobee only include information from indoor movement using sensors, limiting the analysis to indoor behaviors. Integration of these data with other IoT devices such as smartwatches, cell phones, and fitness trackers would provide more comprehensive insights into population-level behaviors.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The Tukey post hoc test compare sleep parameters and time spent at home with respect to different weekday pairs (Table S1); descriptive analysis of seasonal weekdays subgroups for sleep parameters and time spent at home (Table S2); evaluating the homogeneity of variances in different weekdays for all season groups using the Levene test (Table S3).
[DOCX File , 78 KB - mhealth_v10i4e28811_app1.docx]

References

1. Buysse DJ. Sleep health: can we define it? Does it matter? Sleep 2014;37(1):9-17 [FREE Full text] [doi: 10.5665/sleep.3298] [Medline: 24470692]
2. Magnavita N, Garbarino S. Sleep, health and wellness at work: a scoping review. Int J Environ Res Public Health 2017;14(11):1347 [FREE Full text] [doi: 10.3390/ijerph14111347] [Medline: 29113118]
3. Watson S, Cherney K. The effects of sleep deprivation on your body. Healthline. 2021. URL: https://www.healthline.com/health/sleep-deprivation/effects-on-body#Digestive-system [accessed 2021-03-15]
4. Ciman M, Wac K. Smartphones as sleep duration sensors: validation of the iSenseSleep algorithm. JMIR Mhealth Uhealth 2019;7(5):e11930 [FREE Full text] [doi: 10.2196/11930] [Medline: 31115341]
5. Thomée S, Härenstam A, Hagberg M. Mobile phone use and stress, sleep disturbances, and symptoms of depression among young adults--a prospective cohort study. BMC Public Health 2011;11:66 [FREE Full text] [doi: 10.1186/1471-2458-11-66] [Medline: 21281471]
6. Itani O, Jike M, Watanabe N, Kaneita Y. Short sleep duration and health outcomes: a systematic review, meta-analysis, and meta-regression. Sleep Med 2017;32:246-256. [doi: 10.1016/j.sleep.2016.08.006] [Medline: 27743803]
7. Chaput JP, Carson V, Gray CE, Tremblay MS. Importance of all movement behaviors in a 24 hour period for overall health. Int J Environ Res Public Health 2014;11(12):12575-12581 [FREE Full text] [doi: 10.3390/ijerph111212575] [Medline: 25485978]
8. Are Canadian adults getting enough sleep? Public Health Agency of Canada. 2019. URL: https://www.canada.ca/en/public-health/services/publications/healthy-living/canadian-adults-getting-enough-sleep-infographic.html [accessed 2020-12-28]
9. Robbins R, Affouf M, Seixas A, Beaugris L, Avirappattu G, Jean-Louis G. Four-year trends in sleep duration and quality: a longitudinal study using data from a commercially available sleep tracker. J Med Internet Res 2020;22(2):e14735 [FREE Full text] [doi: 10.2196/14735] [Medline: 32078573]
10. Moore SA, Faulkner G, Rhodes RE, Brussoni M, Chulak-Bozzer T, Ferguson LJ, et al. Impact of the COVID-19 virus outbreak on movement and play behaviours of Canadian children and youth: a national survey. Int J Behav Nutr Phys Act 2020;17(1):85 [FREE Full text] [doi: 10.1186/s12966-020-00987-8] [Medline: 32631350]
11. Chaput JP. The integration of pediatric sleep health into public health in Canada. Sleep Med 2019;56:4-8. [doi: 10.1016/j.sleep.2018.06.009] [Medline: 30076116]
12. Saeb S, Cybulski TR, Schueller SM, Kording KP, Mohr DC. Scalable passive sleep monitoring using mobile phones: opportunities and obstacles. J Med Internet Res 2017;19(4):e118 [FREE Full text] [doi: 10.2196/jmir.6821] [Medline: 28420605]
13. Williams JA, Cook DJ. Forecasting behavior in smart homes based on sleep and wake patterns. Technol Health Care 2017;25(1):89-110 [FREE Full text] [doi: 10.3233/THC-161255] [Medline: 27689555]
14. Boger J, Young V, Hooey J, Jiancaro T, Mihailidis A. Zero effort technologies: considerations, challenges, and use in health, wellness, and rehabilitation: synthesis lectures on assistive, rehabilitative, and health-preserving technologies. 2nd edition. San Rafael, CA: Morgan & Claypool Publishers; 2018.
15. Internet of things: new promises for persons with disabilities. G3ict. 2015. URL: https://g3ict.org/upload/publication/internet-of-things-new-promises-for-persons-with-disabilities/IoT_new-promises-for-PWD.pdf [accessed 2021-09-15]

16. Gutierrez LJ, Rabbani K, Ajayi OJ, Gebresilassie SK, Rafferty J, Castro LA, et al. Internet of things for mental health: open issues in data acquisition, self-organization, service level agreement, and identity management. Int J Environ Res Public Health 2021;18(3):1327 [FREE Full text] [doi: 10.3390/ijerph18031327] [Medline: 33535714]
17. Sahu KS, Oetomo A, Morita PP. Enabling remote patient monitoring through the use of smart thermostat data in canada: exploratory study. JMIR Mhealth Uhealth 2020;8(11):e21016 [FREE Full text] [doi: 10.2196/21016] [Medline: 33216001]
18. Jalali N, Sahu KS, Oetomo A, Morita PP. Understanding user behavior through the use of unsupervised anomaly detection: proof of concept using internet of things smart home thermostat data for improving public health surveillance. JMIR Mhealth Uhealth 2020;8(11):e21209 [FREE Full text] [doi: 10.2196/21209] [Medline: 33185562]
19. Kramer O. Scikit-learn. In: Kramer O, editor. Machine learning for evolution strategies. Cham, the Netherlands: Springer International Publishing; 2016:45-53.
20. Kleiminger W, Mattern F, Santini S. Predicting household occupancy for smart heating control: a comparative performance analysis of state-of-the-art approaches. Energy Build 2014;85:493-505. [doi: 10.1016/j.enbuild.2014.09.046]
21. Huchuk B, Sanner S, O'Brien W. Comparison of machine learning models for occupancy prediction in residential buildings using connected thermostat data. Build Environ 2019;160:106177. [doi: 10.1016/j.buildenv.2019.106177]
22. Reynolds D. Gaussian mixture models. In: Li SZ, Jain AK, editors. Encyclopedia of Biometrics. Boston, MA: Springer; 2015:827-832.
23. Larson MG. Analysis of variance. Circulation 2008;117(1):115-121. [doi: 10.1161/CIRCULATIONAHA.107.654335] [Medline: 18172051]
24. Sow MT. J Int Bus Econ 2014;2(4):6. [doi: 10.15640/jibe.v2n4a6]
25. Lix LM, Keselman JC, Keselman HJ. Consequences of assumption violations revisited: a quantitative review of alternatives to the one-way analysis of variance F test. Rev Educ Res 1996;66(4):579-619. [doi: 10.3102/00346543066004579]
26. 3. When do the seasons start? National Research Council Canada. 2020. URL: https://nrc.canada.ca/en/certifications-evaluations-standards/canadas-official-time/3-when-do-seasons-start [accessed 2020-12-17]
27. Gastwirth JL, Gel YR, Miao W. The impact of Levene's test of equality of variances on statistical theory and practice. Stat Sci 2009;24(3):343-360. [doi: 10.1214/09-sts301]
28. Hayes TL, Riley T, Pavel M, Kaye JA. Estimation of rest-activity patterns using motion sensors. Annu Int Conf IEEE Eng Med Biol Soc 2010;2010:2147-2150 [FREE Full text] [doi: 10.1109/IEMBS.2010.5628022] [Medline: 21097221]
29. Kim HY. Statistical notes for clinical researchers: two-way analysis of variance (ANOVA)-exploring possible interaction between factors. Restor Dent Endod 2014;39(2):143-147 [FREE Full text] [doi: 10.5395/rde.2014.39.2.143] [Medline: 24790929]
30. Favaretto M, De Clercq E, Elger BS. Big data and discrimination: perils, promises and solutions. A systematic review. J Big Data 2019;6(1):12. [doi: 10.1186/s40537-019-0177-4]
31. Basner M, Fomberstein KM, Razavi FM, Banks S, William JH, Rosa RR, et al. American time use survey: sleep time and its relationship to waking activities. Sleep 2007;30(9):1085-1095 [FREE Full text] [doi: 10.1093/sleep/30.9.1085] [Medline: 17910380]
32. Chaput JP, Janssen I. Sleep duration estimates of Canadian children and adolescents. J Sleep Res 2016;25(5):541-548 [FREE Full text] [doi: 10.1111/jsr.12410] [Medline: 27027988]
33. Mumu SJ, Fahey PP, Ali L, Rahman AK, Merom D. Seasonal variations in physical activity domains among rural and urban Bangladeshis using a culturally relevant past year physical activity questionnaire (PYPAQ). J Environ Public Health 2019;2019:2375474 [FREE Full text] [doi: 10.1155/2019/2375474] [Medline: 31737080]
34. Yoshimura E, Tajiri E, Hatamoto Y, Tanaka S. Changes in season affect body weight, physical activity, food intake, and sleep in female college students: a preliminary study. Int J Environ Res Public Health 2020;17(23):8713 [FREE Full text] [doi: 10.3390/ijerph17238713] [Medline: 33255205]
35. Watson NF, Badr MS, Belenky G, Bliwise DL, Buxton OM, Buysse D, et al. Recommended amount of sleep for a healthy adult: a joint consensus statement of the American academy of sleep medicine and sleep research society. Sleep 2015;38(6):843-844 [FREE Full text] [doi: 10.5665/sleep.4716] [Medline: 26039963]
36. Li W, Bertisch SM, Mostofsky E, Vgontzas A, Mittleman MA. Associations of daily weather and ambient air pollution with objectively assessed sleep duration and fragmentation: a prospective cohort study. Sleep Med 2020;75:181-187 [FREE Full text] [doi: 10.1016/j.sleep.2020.06.029] [Medline: 32858358]
37. Hashizaki M, Nakajima H, Shiga T, Tsutsumi M, Kume K. A longitudinal large-scale objective sleep data analysis revealed a seasonal sleep variation in the Japanese population. Chronobiol Int 2018;35(7):933-945. [doi: 10.1080/07420528.2018.1443118] [Medline: 29589960]
38. Åkerstedt T, Ghilotti F, Grotta A, Zhao H, Adami HO, Trolle-Lagerros Y, et al. Sleep duration and mortality - does weekend sleep matter? J Sleep Res 2019;28(1):e12712 [FREE Full text] [doi: 10.1111/jsr.12712] [Medline: 29790200]
39. Hashizaki M, Nakajima H, Kume K. Monitoring of weekly sleep pattern variations at home with a contactless biomotion sensor. Sensors (Basel) 2015;15(8):18950-18964 [FREE Full text] [doi: 10.3390/s150818950] [Medline: 26247948]
40. Zhang R, Tomasi D, Shokri-Kojori E, Wiers CE, Wang GJ, Volkow ND. Sleep inconsistency between weekends and weekdays is associated with changes in brain function during task and rest. Sleep 2020;43(10):zsaa076 [FREE Full text] [doi: 10.1093/sleep/zsaa076] [Medline: 32333599]

41. Bhat CR, Misra R. Discretionary activity time allocation of individuals between in-home and out-of-home and between weekdays and weekends. Transportation 1999;26(2):193-229. [doi: 10.1023/A:1005192230485]
42. Khajehzadeh I, Vale B. How New Zealanders distribute their daily time between home indoors, home outdoors and out of home. Kōtuitui 2017;12(1):17-31. [doi: 10.1080/1177083x.2016.1187636]
43. Hjorth MF, Chaput JP, Michaelsen K, Astrup A, Tetens I, Sjödin A. Seasonal variation in objectively measured physical activity, sedentary time, cardio-respiratory fitness and sleep duration among 8-11 year-old Danish children: a repeated-measures study. BMC Public Health 2013;13:808 [FREE Full text] [doi: 10.1186/1471-2458-13-808] [Medline: 24010811]
44. Rich C, Griffiths LJ, Dezateux C. Seasonal variation in accelerometer-determined sedentary behaviour and physical activity in children: a review. Int J Behav Nutr Phys Act 2012;9:49 [FREE Full text] [doi: 10.1186/1479-5868-9-49] [Medline: 22546178]
45. O'Connell S, Griffiths P, Clemes S. Seasonal variation in physical activity, sedentary behaviour and sleep in UK adults. J Sci Med Sport 2012;15(1):S19. [doi: 10.1016/j.jsams.2012.11.046]
46. Bittman M. Sunday working and family time. Labour Ind 2005;16(1):59-81. [doi: 10.1080/10301763.2005.10722031]
47. de Bell S, White M, Griffiths A, Darlow A, Taylor T, Wheeler B, et al. Spending time in the garden is positively associated with health and wellbeing: results from a national survey in England. Landsc Urban Plan 2020;200:103836. [doi: 10.1016/j.landurbplan.2020.103836]
48. White MP, Alcock I, Grellier J, Wheeler BW, Hartig T, Warber SL, et al. Spending at least 120 minutes a week in nature is associated with good health and wellbeing. Sci Rep 2019;9(1):7730 [FREE Full text] [doi: 10.1038/s41598-019-44097-3] [Medline: 31197192]
49. Twohig-Bennett C, Jones A. The health benefits of the great outdoors: a systematic review and meta-analysis of greenspace exposure and health outcomes. Environ Res 2018;166:628-637 [FREE Full text] [doi: 10.1016/j.envres.2018.06.030] [Medline: 29982151]
50. Walden S. The "Indoor Generation" and the health risks of spending more time inside. USA Today. 2018. URL: https://www.usatoday.com/story/sponsor-story/velux/2018/05/15/indoor-generation-and-health-risks-spending-more-time-inside/610289002/ [accessed 2020-09-22]
51. Manferdelli G, La Torre A, Codella R. Outdoor physical activity bears multiple benefits to health and society. J Sports Med Phys Fitness 2019;59(5):868-879. [doi: 10.23736/S0022-4707.18.08771-6] [Medline: 30650943]
52. An R, Zhang S, Ji M, Guan C. Impact of ambient air pollution on physical activity among adults: a systematic review and meta-analysis. Perspect Public Health 2018;138(2):111-121. [doi: 10.1177/1757913917726567] [Medline: 28829249]
53. Cabeza de Baca T, Chayama KL, Redline S, Slopen N, Matsushita F, Prather AA, et al. Sleep debt: the impact of weekday sleep deprivation on cardiovascular health in older women. Sleep 2019;42(10):zsz149 [FREE Full text] [doi: 10.1093/sleep/zsz149] [Medline: 31361895]
54. Plasqui G, Westerterp KR. Seasonal variation in total energy expenditure and physical activity in Dutch young adults. Obes Res 2004;12(4):688-694 [FREE Full text] [doi: 10.1038/oby.2004.80] [Medline: 15090638]
55. Matz CJ, Stieb DM, Davis K, Egyed M, Rose A, Chou B, et al. Effects of age, season, gender and urban-rural status on time-activity: CanadianHuman Activity Pattern Survey 2 (CHAPS 2). Int J Environ Res Public Health 2014;11(2):2108-2124 [FREE Full text] [doi: 10.3390/ijerph110202108] [Medline: 24557523]
56. Farrow A, Taylor H, Golding J. Time spent in the home by different family members. Environ Technol 1997;18(6):605-613. [doi: 10.1080/09593331808616578]
57. Lauderdale DS, Knutson KL, Yan LL, Liu K, Rathouz PJ. Self-reported and measured sleep duration: how similar are they? Epidemiology 2008;19(6):838-845 [FREE Full text] [doi: 10.1097/EDE.0b013e318187a7b0] [Medline: 18854708]
58. Carney CE, Buysse DJ, Ancoli-Israel S, Edinger JD, Krystal AD, Lichstein KL, et al. The consensus sleep diary: standardizing prospective sleep self-monitoring. Sleep 2012;35(2):287-302 [FREE Full text] [doi: 10.5665/sleep.1642] [Medline: 22294820]
59. van Hees VT, Sabia S, Jones SE, Wood AR, Anderson KN, Kivimäki M, et al. Estimating sleep parameters using an accelerometer without sleep diary. Sci Rep 2018;8(1):12975 [FREE Full text] [doi: 10.1038/s41598-018-31266-z] [Medline: 30154500]
60. de Zambotti M, Cellini N, Goldstone A, Colrain IM, Baker FC. Wearable sleep technology in clinical and research settings. Med Sci Sports Exerc 2019;51(7):1538-1557 [FREE Full text] [doi: 10.1249/MSS.0000000000001947] [Medline: 30789439]
61. Scott H, Biello SM, Woods HC. Social media use and adolescent sleep patterns: cross-sectional findings from the UK millennium cohort study. BMJ Open 2019;9(9):e031161 [FREE Full text] [doi: 10.1136/bmjopen-2019-031161] [Medline: 31641035]
62. Sahu KS, Majowicz SE, Dubin JA, Morita PP. NextGen public health surveillance and the internet of things (IoT). Front Public Health 2021;9:756675 [FREE Full text] [doi: 10.3389/fpubh.2021.756675] [Medline: 34926381]
63. Matar G, Lina MJ, Carrier J, Kaddoum G. Unobtrusive sleep monitoring using cardiac, breathing and movements activities: an exhaustive review. IEEE Access 2018;6:45129-45152. [doi: 10.1109/access.2018.2865487]
64. Berger L, Berger N, Bosetti V, Gilboa I, Hansen LP, Jarvis C, et al. Rational policymaking during a pandemic. Proc Natl Acad Sci U S A 2021;118(4):e2012704118 [FREE Full text] [doi: 10.1073/pnas.2012704118] [Medline: 33472971]

Abbreviations

ANOVA: analysis of variance
DYD: Donate Your Data
IoT: Internet of Things
OR: odds ratio

Edited by L Buis; submitted 16.03.21; peer-reviewed by D Istrat, S Mukherjee; comments to author 16.04.21; revised version received 21.04.21; accepted 03.02.22; published 01.04.22.

Please cite as:
Jalali N, Sahu KS, Oetomo A, Morita PP
Usability of Smart Home Thermostat to Evaluate the Impact of Weekdays and Seasons on Sleep Patterns and Indoor Stay: Observational Study
JMIR Mhealth Uhealth 2022;10(4):e28811
URL: https://mhealth.jmir.org/2022/4/e28811
doi:10.2196/28811
PMID:35363147

Original Paper

An Unstructured Supplementary Service Data–Based mHealth App Providing On-Demand Sexual Reproductive Health Information for Adolescents in Kibra, Kenya: Randomized Controlled Trial

Paul Macharia[1*], MSc; Antoni Pérez-Navarro[1*], PhD; Betsy Sambai[2*], MSc; Irene Inwani[2*], MD, MPH; John Kinuthia[2*], MD, MPH; Ruth Nduati[3*], MD, MPH; Carme Carrion[1*], PhD

[1]Universitat Oberta de Catalunya, Barcelona, Spain

[2]Kenyatta National Hospital, Nairobi, Kenya

[3]University of Nairobi, Nairobi, Kenya

[*]all authors contributed equally

Corresponding Author:
Paul Macharia, MSc
Universitat Oberta de Catalunya
Rambla del Poblenou, 156
Barcelona, 08018
Spain
Phone: 34 619963691
Email: paulmachariah@gmail.com

Abstract

Background: Adolescents transitioning from childhood to adulthood need to be equipped with sexual reproductive health (SRH) knowledge, skills, attitudes, and values that empower them. Accessible, reliable, appropriate, and friendly information can be provided through mobile phone–based health interventions.

Objective: This study aims to investigate the effectiveness and impact of an Unstructured Supplementary Service Data (USSD)–based app in increasing adolescents' knowledge about contraceptives, gender-based stereotypes, sexually transmitted infections (STIs), abstinence, and perceived vulnerability, and helping adolescents make informed decisions about their SRH.

Methods: A randomized controlled trial (RCT) methodology was applied to investigate the potential of a USSD-based app for providing on-demand SRH information. To be eligible, adolescents aged 15 to 19 years residing in Kibra, Kenya, had to have access to a phone and be available for the 3-month follow-up visit. Participants were randomly assigned to the intervention (n=146) and control (n=154) groups using sequentially numbered, opaque, sealed envelopes. The primary outcome was improved SRH knowledge. The secondary outcome was improved decision-making on SRH. The outcomes were measured using validated tools on adolescent SRH and user perceptions during the follow-up visit. A paired sample *t* test was used to compare the changes in knowledge scores in both groups. The control group did not receive any SRH information.

Results: During the RCT, 54.9% (62/109) of adolescents used the USSD-based app at least once. The mean age by randomization group was 17.3 (SD 1.23) years for the control group and 17.3 (SD 1.12) years for the intervention group. There was a statistically significant difference in the total knowledge scores in the intervention group (mean 10.770, SD 2.012) compared with the control group (mean 10.170, SD 2.412) conditions (t_{179}=2.197; *P*=.03). There was a significant difference in abstinence (*P*=.01) and contraceptive use (*P*=.06). Of the individuals who used the app, all participants felt the information received could improve decision-making regarding SRH. Information on STIs was of particular interest, with 27% (20/62) of the adolescents seeking information in this area, of whom 55% (11/20) were female. In relation to improved decision-making, 21.6% (29/134) of responses showed the adolescents were able to identify STIs and were likely to seek treatment; 51.7% (15/29) of these were female. Ease of use was the most important feature of the app for 28.3% (54/191) of the responses.

Conclusions: Adolescents require accurate and up-to-date SRH information to guide their decision-making and improve health outcomes. As adolescents already use mobile phones in their day-to-day lives, apps provide an ideal platform for this information. A USSD-based app could be an appropriate tool for increasing SRH knowledge among adolescents in low-resource settings.

Adolescents in the study valued the information provided because it helped them identify SRH topics on which they needed more information.

Trial Registration: Pan African Clinical Trial Registry PACTR202204774993198; https://pactr.samrc.ac.za/TrialDisplay.aspx?TrialID=22623

(JMIR Mhealth Uhealth 2022;10(4):e31233) doi:10.2196/31233

KEYWORDS

adolescents; sexual reproductive health; mobile phones; randomized controlled trial

Introduction

Background

The World Health Organization (WHO) has stated that universal access to quality sexual reproductive health (SRH) services is essential for sustainable development and global realization of health and human rights [1]. The United Nations has also made a commitment to ensure "universal access to sexual health and reproductive health-care services, including family planning, information and education" [2]. As SRH rights are fundamental to humanity's well-being, the provision of evidence-based SRH interventions will secure lifelong positive impacts on health benefits and outcomes [3].

Adolescents transitioning from childhood to adulthood must be equipped with SRH knowledge, skills, attitudes, and values that empower them to develop successful sexual relationships. Several approaches, including comprehensive sexuality education and curriculum-based approaches, have been used to teach adolescents different aspects of sexuality [4]. Health interventions can be expanded to settings that adolescents engage in, beyond family and health care facilities [5]. Adolescents can be provided with high-impact, easily accessible, and reliable health information that is crucial for improving their reproductive health [6]. Content should be adolescent-friendly, appropriate to their SRH needs, swiftly provided, and not overwhelming [7].

In low- and middle-income countries, the exponential growth of mobile-based technologies has provided opportunities for the adoption of mobile health (mHealth) apps. The WHO identifies many mobile phone technologies that can be used to improve health outcomes in low- and middle-income countries, including SMS text messaging [8]. As research has shown, modes of information delivery and content must vary according to audience, appealing to different users in different ways [9]. In resource-limited settings, for instance, technology-based interventions have proven to be an effective way of providing health information [10].

Using evidence-based content to deliver adolescent SRH information on mobile phones has the potential to impact behaviors and improve health outcomes [11]. Several mobile phone–based interventions providing adolescent SRH services and their impact have been well-documented [12]. Research has shown that mHealth interventions have the potential to engage adolescents across sociodemographic settings and increase their knowledge and awareness [13]. Such interventions appeal to adolescents and, therefore, can mitigate the barriers to access associated with the delivery of adolescent SRH information at health care facilities [14].

Mobile phone–based health interventions are an increasingly feasible way to connect adolescents with SRH information and services in low-resource settings. Research has shown that interventions have been able to provide adolescents with knowledge that can lead to behavior change and improved health outcomes [15]. Mobile phone–based interventions can be tailored to each adolescent's context and provide individualized and effective services [16]. To improve such interventions, it is important to document and review system interaction data to inform design and delivery improvements, thereby making mHealth apps more effective [17]. There are often concerns about privacy when using mobile phone apps, which must be considered during the app development process [15]. Unstructured Supplementary Service Data (USSD)-based mobile phone technology has been found to be a user-friendly, convenient, and confidential method for adolescent users to access SRH information [18,19].

Objective

This study investigates the potential of a USSD-based app for providing on-demand SRH information to adolescents in the resource-limited setting of Kibra, Nairobi County, Kenya. The aim of this study is to determine the effectiveness and impact of a USSD-based mobile phone app in (1) increasing adolescents' knowledge about contraceptives, gender-based stereotypes, sexually transmitted infections (STIs), abstinence, and perceived vulnerability and (2) helping adolescents make informed decisions about their SRH.

Methods

Ethics Approval

The study protocol was reviewed and approved by the Kenyatta National Hospital University of Nairobi Ethics Review Committee in March 2019 (reference number P707/10/2018).

Intervention Design

The intervention design was based on the health belief model, a behavior change framework intended to increase knowledge that can inform actions to reduce health risks [20]. Through a randomized controlled trial (RCT), a USSD-based app was evaluated on its ability to influence adolescents' knowledge, attitudes, and practices related to SRH health awareness. Content provided in the app was based on validated adolescent sexual health information (Multimedia Appendix 1) created through Avert's Young Voices, a project that developed materials and content on adolescent sexual health through a co-creation

process with adolescents from South Africa, Lesotho, Zambia, Zimbabwe, and Malawi [21].

Sample Size

The study enrolled 300 adolescents: 146 (48.7%) randomized to the intervention group and 154 (51.3%) to the control group. It is estimated that around 8% of adolescents aged 15 to 19 years in Kenya access SRH information [22]. A minimum sample size of 226 adolescents was required to attain a 95% CI. Thus, the sample size of 300 adolescents had 74 more participants than the minimum sample size. The additional participants ensured that the sample strength would be maintained, even with loss at follow-up.

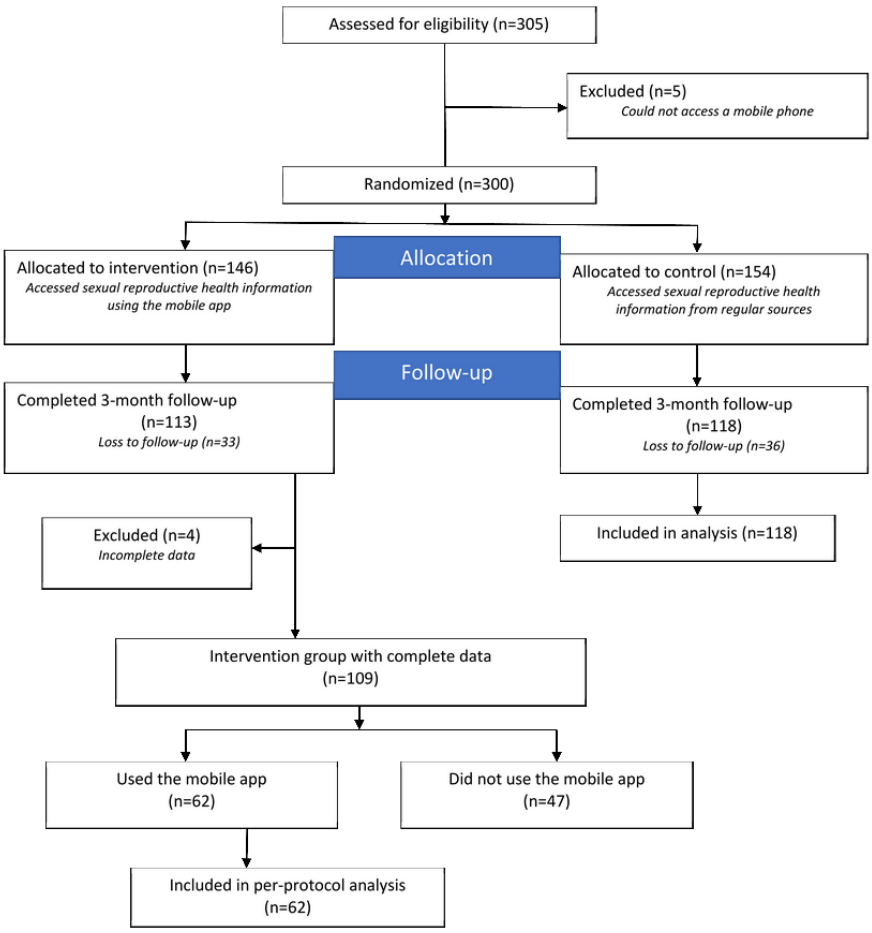
The study used sequentially numbered, opaque sealed envelopes—an affordable and effective method for randomizing participants [23]. Having passed screening for eligibility, the adolescents picked a sealed envelope from a box. Each envelope contained a randomization group and an assigned participant ID number. A randomization list was generated using a web-based tool [24].

Recruitment

The study population consisted of adolescents aged 15 to 19 years residing in the Kibra suburb of Nairobi City County, Kenya. Kibra consists of 12 villages, with both formal and informal settlements. The informal settlements house approximately 2.5 million residents. Participants were mobilized from 12 villages; community mobilizers approached potential participants at social halls, sports events, and other social activities that attracted adolescents aged 15 to 19 years. The study procedures were explained individually or to small groups of 3 to 5 adolescents using a study recruitment script (Multimedia Appendix 2). Adolescents interested in the study were referred to the study venue.

Efforts were made to distribute enrollment across all villages, as there are intervillage ethnic differences. Enrollment numbers were monitored by village and randomization groups during enrollment to ensure equitable distribution, providing an improved representation of adolescent SRH needs and awareness across the area. Ethnicity data were not collected because of the stigma associated with issues or discussions on ethnicity in the study site setting. We distributed enrollment across all the villages, ensuring a true representation of Kibra. In 2 cases, the enrollment team moved the study site to a social hall near a particular village to make it easier for local adolescents to participate. Figure 1 shows the enrollment and follow-up processes.

Figure 1. Study participant flowchart.



Inclusion and Exclusion Criteria

To participate in the study, adolescents should be aged between 15 and 19 years, live in Kibra, and be able to access a mobile phone. Participants aged 15 to 17 years signed an assent form after assenting to the study procedures; those aged ≥18 years were required to sign a consent form.

Adolescents aged <18 years should be accompanied by a parent or guardian, should get the permission from the parent or guardian, and should provide their assent. As the study presented minimal risk, the study team requested a waiver of parental permission for adolescents aged <18 years who were unaccompanied. In this setting, there are cultural challenges related to discussions with parents regarding adolescent SRH. Parents or guardians in Kibra, as in many settings, may not be involved in or fully aware of their adolescents' SRH information needs. If the study opted to secure parental permission for adolescents participating in the study, this may have required disclosure of the participants' SRH information needs, potentially leading to an elevated risk of harm or prevention of participation.

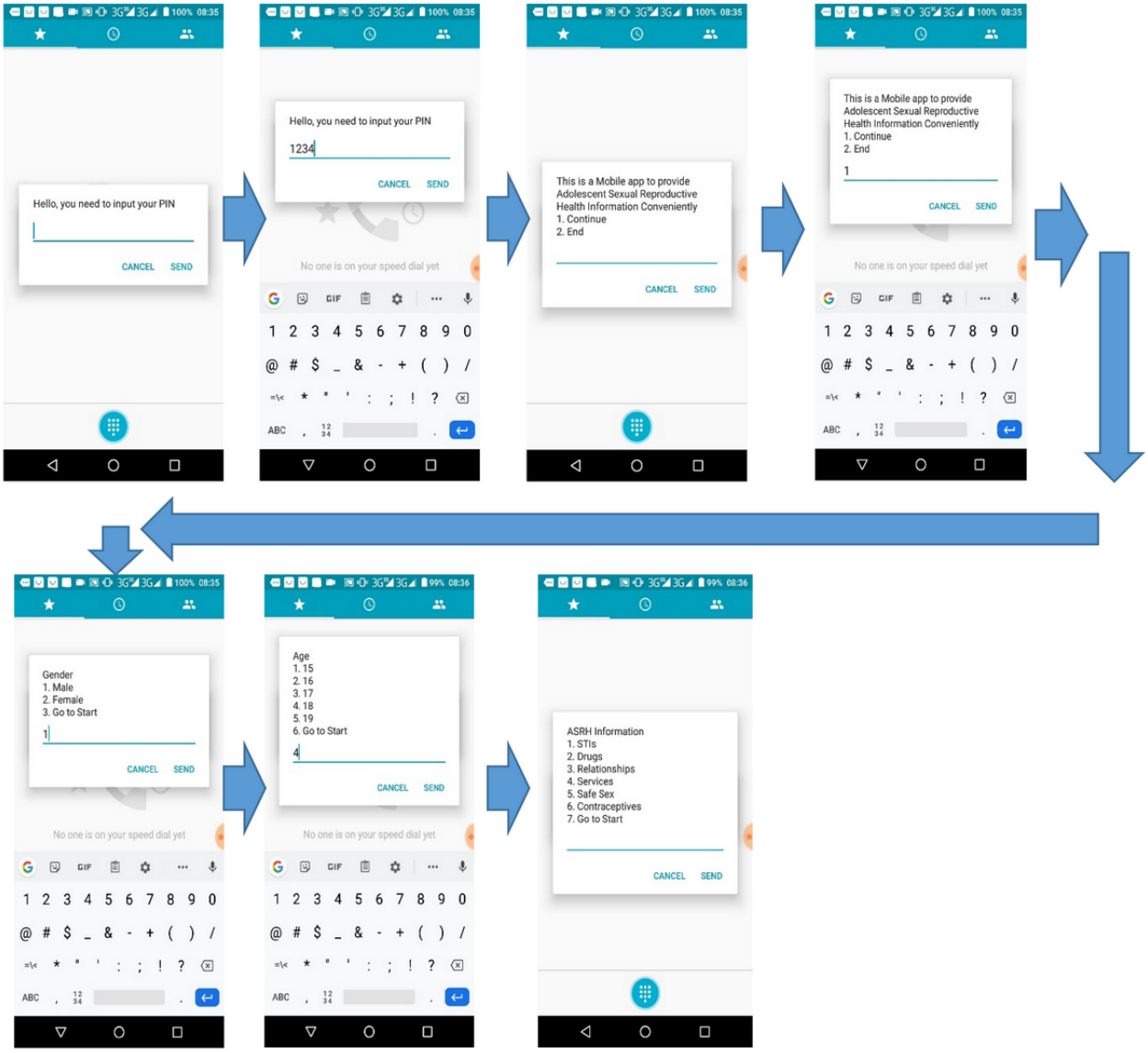
Intervention Implementation and Data Collection

At the study venue, all potential participants were provided with further details of the study, eligibility criteria, and study procedures. On the basis of their age, an approved informed consent or assent form was provided in either English or Swahili. Potential participants were given time to ask questions, and after these were addressed, the study staff verified whether the potential participants were still interested. After the study procedures were explained in detail, participants signed a consent or assent form. Study staff then signed and dated the consent or assent forms, and participants were provided with a copy, if desired. For adolescents in the control group, no SRH information was provided. It was assumed that these adolescents would get information from their regular sources, including their parents, peers, or seminars held by nongovernmental organizations in Kibra.

The *Evaluation of Knowledge of SRH Information* (Multimedia Appendix 3) and *Use and Perception of the Mobile Phone App* (Multimedia Appendix 4) questionnaires were administered to adolescents by study staff. Open Data Kit was used to administer the 2 questionnaires. This mobile app enables a survey to be administered through a smartphone, question by question, in an easy-to-use, user-friendly interface. The *Evaluation of Knowledge of SRH Information* questionnaire was administered at enrollment and follow-up visits. The *Use and Perception of the Mobile Phone App* questionnaire was only administered to adolescents in the intervention group who had used the app at least once.

For the follow-up visit, community mobilizers called each participant, requesting them to visit the study site. Follow-up interviews were scheduled based on adolescents' availability. A USSD app offering validated SRH information accessible on both feature phones and smartphones was availed to the adolescents in the intervention group. To visualize the USSD app, Figure 2 shows the layout of the app and its interactive screens. A user-centered design approach was used in the design and development of the USSD app [18].

Figure 2. Connecting to the Unstructured Supplementary Service Data (USSD) app. STI: sexually transmitted infection.



Sexual Reproductive Health Knowledge Score

To evaluate intervention outcomes, the *Evaluation of Knowledge of SRH Information* questionnaire (Multimedia Appendix 3) based on Monitoring and Evaluation to Assess and Use Results evaluation indicators on adolescent SRH was administered during the 3-month follow-up visit [25]. The knowledge scores calculated from these questionnaires were used to evaluate awareness.

The questions required yes or no responses. Each correct response earned one point and the wrong answer scored zero. Questions were posed to adolescents in both the intervention and control groups at enrollment and follow-up visits. The knowledge score data were analyzed as an aggregate score and by each subsection of the *Evaluation of Knowledge of SRH Information* questionnaire (Multimedia Appendix 3). Only the questions in section 2 of the questionnaire were used to evaluate knowledge scores. All estimates were adjusted for age, sex, access to phones, and level of education.

Use and Perceptions of the Mobile Phone App

The *Use and Perception of the Mobile Phone App* questionnaire (Multimedia Appendix 4) was administered to participants who had used the app at least once during the 3-month period. This questionnaire is based on WHO-validated instruments intended to be used by investigators studying SRH among adolescents [26] and assesses potentially improved decision-making. Knowledge of contraceptive use, STIs, and abstinence was also assessed. This study paid particular attention to detailed descriptions of recent encounters to determine the intervention's impact on improved awareness of SRH information.

Statistical Analysis

Overview

A paired sample *t* test was conducted to compare the knowledge score data. The *t* test attempted to show if there were differences in knowledge scores between the intervention and control groups at the 3-month follow-up visit. We also attempted to assess changes in attitude toward contraceptives, gender role

stereotypes, abstinence, and perceived vulnerability. The data were analyzed using R software (R Core Team).

For the *Use and Perception of the Mobile Phone App* questionnaire data, descriptive statistics were used to assess adolescents' knowledge, awareness, and potentially improved decision-making in relation to SRH. A chi-square test was used to assess any differences by age group, with a *P* value <.05 regarded as significant. Data analysis was performed using the R software (version 3.6.2).

Univariate Analyses

Exploratory data analysis techniques were performed to reveal the distribution structure of the outcome variables and identify outliers or unusually entered values. Statistical analyses were performed using descriptive statistics for continuous (mean and SD) and categorical (frequency and proportion) variables. These tests were performed on each participant's demographics and the *Use and Perception of the Mobile Phone App* data.

Bivariate Analysis

The distribution of the background characteristics of the study groups was compared. To establish baseline equivalence between the intervention and control groups, 2 analytical tests were used. The distribution of categorical variables (proportions) was compared using Pearson chi-square test, whereas the distribution of continuous variables (mean and SD) was compared using an independent *t* test. Bivariate analysis was also performed on participant demographics, and the use and perception of mobile phone app data. A *P* value <.05 was regarded as significant.

Analysis of the Effect of the Intervention

Longitudinal continuous outcome scores were analyzed across time points (baseline and end line) to understand the effect of variations in outcome scores. The 2-tailed paired sample *t* test would compare the means of the intervention and control groups. The continuous outcomes were normally distributed. The threshold for statistical significance for all analyses was set at *P*<.05. This analysis was performed on the knowledge score data.

Results

Overview

In October 2019, 305 adolescents from 12 villages in Kibra were mobilized and screened for eligibility. Owing to lack of access to a mobile phone, 5 adolescents were excluded from the study. Study participants were then randomized to the intervention (154/300, 51.3%) and control (146/300, 48.7%) groups. From late December 2019, 77% (231/300) of the adolescents were successfully followed up—74.7% (109/300) from the intervention group and 76.6% (118/300) from the control group. As the app was not used at least once, 47 participants, together with another 4 participants with incomplete data in the intervention group, were excluded from the final analysis. Figure 1 shows the enrollment and follow-up stages of the study. The data were analyzed as per the per-protocol analysis.

Background Characteristics of the Study Participants

The distribution of the study participants according to the selected background characteristics indicated a desired comparable result at baseline as shown in Table 1. The mean ages of the participants in the control group (17.29, SD 1.23 years) and the intervention group (17.27, SD 1.12 years) were statistically comparable (*P*=.94). There were no significant differences in participant demographics. However, there was a statistically significant difference in the distribution of sex (*P*=.03) by the study enrollment group.

Table 1. Characteristics of the study participants (N=300).

Variables	Total (n=180), n (%)	Intervention (n=62), n (%)	Control (n=118), n (%)	*P* value
Sex				.03
Male	67 (37.2)	30 (48.4)	37 (31.4)	
Female	113 (62.8)	32(51.6)	81 (68.6)	
Phone ownership				.64
Adolescent	81 (45)	30 (48.4)	51 (43.2)	
Parent or guardian	90 (50)	30 (48.4)	60 (50.8)	
Other	9 (5)	2 (3.2)	7 (5.9)	
Highest level of education				.97
Primary	23 (12.8)	8 (12.9)	15 (12.7)	
Secondary and above	157 (87.2)	54 (87.1)	103 (87.3)	

SRH Knowledge Score

Participants' responses were analyzed by attitude toward contraceptives, gender role stereotypes, abstinence, and perceived vulnerability to negative SRH outcomes. Knowledge scores were also analyzed as aggregated data. A paired sample *t* test analysis of the relationship between the knowledge score and the use of the mobile app was performed using R software.

Table 2 presents an analysis of the effect of the intervention on specific indicator scores. The difference in the mean scores between those enrolled in the intervention group compared with those in the control group showed statistical significance in the

total knowledge scores. The overall mean change in total scores in the intervention group was 0.5 (*P*=.02) compared with the control group 0.246 (*P*=.24). The *P* value between the 2 groups on the total knowledge scores was .03, which was statistically significant, indicating that the mobile app had an impact on the adolescents' SRH knowledge scores. In the intervention group, the intervention had a statistically significant effect on contraceptive scores (0.355; *P*=.02). The intervention also showed a trend toward statistical significance in abstinence knowledge scores (0.129, *P*=.09).

Table 2. Effects of intervention on overall and specific knowledge scores.

Outcome (knowledge score)	Intervention					Control					Between group, *P* value
	Baseline, mean (SD); 95% CI	End line, mean (SD); 95% CI	Difference in scores, mean (SD); 95% CI	Effect sizes	Within group, *P* value	Baseline, mean (SD); 95% CI	End line, mean (SD); 95% CI	Difference in score, mean (SD); 95% CI	Effect sizes	Within group, *P* value	
Contraceptives	3.613 (1.107); 3 to 4	3.968 (0.887); 4 to 5	0.355 (1.147); 0.064 to 0.646	0.309	.02	3.602 (1.039); 3 to 4	3.678 (1.183); 3 to 5	0.076 (1.235); −0.148 to 0.301	0.062	.5	.06
Vulnerability	2.000 (0.768); 1.25 to 3	2.032 (0.768); 2 to 3	0.0323 (0.829); –0.178 to 0.243	0.038	.76	1.856 (0.860); 1 to 2.75	1.941 (0.798); 1 to 2.75	0.085 (0.939); −0.086 to 0.256	0.090	.33	.32
Gender stereotype	3.097 (0.987); 3 to 4	3.081 (1.060); 3 to 4	−0.016 (0.757); −0.208 to 0.176	0.021	.87	2.890 (1.160); 2 to 4	2.881 (1.126); 2 to 4	−0.008 (1.121); −0.213 to 0.196	0.008	.94	.88
Abstinence	1.565 (0.532); 1 to 2	1.694 (0.465); 1 to 2	0.129 (0.586); −0.020 to 0.278	0.220	.09	1.576 (0.576); 1 to 2	1.669 (0.539); 1 to 2	0.129 (0.569); −0.011 to 0.197	0.163	.08	.01
Total knowledge score	10.270 (2.050); 9 to 12	10.770 (2.012); 10 to 12	0.5 (1.576); 0.099 to 0.900	0.317	.02	9.924 (2.227); 8.25 to 12	10.170 (2.412); 9 to 12	0.246 (2.242); −0.163 to 0.654	0.109	.24	.03

Use and Perceptions of the Mobile Phone App

The use and perceptions questionnaires were used to measure the perceived usefulness of the app. We also aimed to evaluate how the knowledge adolescents received from the app influenced their SRH decision-making. Tables 3 and 4 show the descriptive statistics of our evaluation. The tables show the responses from each adolescent who had used the mobile app at least once in 3 months. The questions addressed topics of interest, the perceived usefulness of information, and the mobile app features the users appreciated. The information in Table 3 is stratified by age—adolescents aged <18 years and those ≥18 years. Table 4 is stratified by gender.

Information about STIs was of great interest to the participants, with 26.7% (20/75) of the responses by users seeking information on this subject the last time they used the app. Adolescent girl participants had a higher interest in STIs, with 55% (11/20) accessing this information. Most participants (56/62, 90.8%) found the information provided in the app to have adequately answered their questions or met their SRH information needs. All the 62 adolescents who used the app felt that the information they received could improve their decision-making on issues relating to SRH. This outcome was similar when data were stratified by age and gender.

The participants reported gaining knowledge from the app on several SRH issues in their responses, including abstinence (53/125, 42.4%), STIs (30/125, 24%), and condom use (22/125, 17.6%). Although only 9.7% (12/125) of the participant's responses showed increased knowledge of contraceptives, 75% (9/12) of these were female, showing a trend toward significance (*P*=.08).

On improved decision-making, 38.1% (51/134) of the adolescent participant's responses show they were able to abstain from sex. Of these responses, 54.9% (28/51) were aged between 15 and 17 years and 52.9% (27/51) were male. The knowledge obtained may have also prompted 26.9% (36/134) of the responses to show use a condom by the adolescent participants during a sexual encounter. Although sex is illegal for ages under 18 years in Kenya, 50% of those who reported deciding to use a condom were aged ≤17 years. Of the participants who used a condom, 52.8% (19/36) were male. Adolescent participants were also able to identify STIs, with 21.6% (29/134) responses reporting that app information guided their decision to seek treatment after identifying an STI; 51.7% (15/29) of these responses were from female participants.

Ease of use was the most important feature of the app for 28.3% (54/191) of the participants' responses, followed by confidentiality at 26.7% (51/191) and high-quality information

at 23.6% (45/191), with 60% (27/45) of the latter being from responses by female participants.

Table 3. Use and perception of the mobile app stratified by age groups (62 participants).

Variable	All, n (%)	Age <18 years, n (%)	Age ≥18 years, n (%)	*P* value
What information did you require when you last used the mobile app?				
STIs[a]	20 (26.7)	8 (40)	12 (60)	.37
Drugs	18 (24)	12 (66.7)	6 (33.3)	.16
Relationship	17 (22.7)	9 (52.9)	8 (47.1)	.81
Sex	12 (16)	4 (33.3)	8 (66.7)	.25
Contraceptives	6 (8)	3 (50)	3 (50)	>.99
Pregnancy	2 (2.7)	0 (0)	2 (100)	.16
What knowledge about SRH[b] issues have you gained?				
Abstinence	53 (42.4)	28 (52.8)	25 (47.2)	.68
STIs	30 (24)	14 (46.7)	16 (53.3)	.72
Condom use	22 (17.6)	12 (54.5)	10 (45.5)	.67
Contraceptives	12 (9.6)	6 (50)	6 (50)	>.99
Drugs	8 (6.4)	5 (62.5)	3 (37.5)	.48
What decision-making was informed by the information you accessed on the mobile app?				
Abstinence	51 (38.1)	28 (54.9)	23 (45.1)	.48
Condom use	36 (26.9)	18 (50)	18 (50)	>.99
STIs	29 (21.6)	11 (37.9)	18 (62.1)	.19
Contraceptives	9 (6.7)	6 (66.7)	3 (33.3)	.32
Drugs	9 (6.7)	6 (66.7)	3 (33.3)	.32
Were the questions you had on SRH answered adequately?				
Yes	56 (90.3)	30 (53.6)	26 (46.4)	>.99
No	6 (9.7)	3 (50)	3 (50)	
Did the information you receive inform better decision-making on SRH matters?				
Yes	62 (100)	33 (53.2)	29 (46.8)	.62
No	0 (0)	0 (0)	0 (0)	
What are the most important features of the mobile phone app?				
Ease of use	54 (28.3)	28 (51.9)	26 (48.1)	.79
Confidentiality	51 (26.7)	26 (51)	25 (49)	.89
Quality of information	45 (23.6)	24 (53.3)	21 (46.7)	.65
Immediate feedback	41 (21.5)	20 (48.8)	21 (51.2)	.88

[a]STI: sexually transmitted infection.

[b]SRH: sexual reproductive health.

Table 4. Use and perception of the mobile app stratified by gender (62 participants).

Variable	All, n (%)	Male, n (%)	Female, n (%)	*P* value
What information did you require when you last used the mobile app?				
STIs[a]	20 (27)	9 (45)	11 (55)	.65
Drugs	18 (24.3)	10 (55.6)	8 (44.4)	.64
Relationships	16 (21.6)	8 (47.1)	9 (52.9)	.81
Sex	12 (16.2)	5 (41.7)	7 (58.3)	.56
Contraceptives	6 (8.1)	1 (16.7)	5 (83.3)	.10
Pregnancy	2 (2.7)	1 (50)	1 (50)	>.99
What knowledge about sexual reproductive health matters have you gained?				
Abstinence	53 (42.7)	26 (49.1)	27 (50.9)	.89
STIs	30 (24.2)	13 (43.3)	17 (56.7)	.47
Condom use	22 (17.7)	13 (50.1)	9 (40.9)	.39
Contraceptives	12 (9.7)	3 (25)	9 (75)	.08
Drugs	7 (5.6)	2 (25)	6 (75)	.16
What better decision-making was informed by the information you accessed on the mobile app?				
Abstinence	51 (38.1)	27 (52.9)	24 (48.1)	.67
Condom use	36 (26.9)	19 (52.8)	17 (47.2)	.74
STIs	29 (21.6)	14 (48.3)	15 (51.7)	.85
Contraceptives	9 (6.7)	4 (44.4)	5 (55.6)	.74
Drugs	9 (6.7)	3 (33.3)	6 (66.7)	.32
Were the questions you had on SRH[b] answered adequately?				
Yes	56 (90.3)	26 (46.4)	30 (53.6)	.61
No	6 (9.7)	4 (33.3)	2 (66.7)	
Did the information you receive inform better decision-making on SRH matters?				
Yes	62 (100)	30 (48.4)	32 (51.6)	.80
No	0 (0)	0	0	
What are the most important features of the mobile phone app?				
Easy to use	54 (28.3)	26 (48.1)	28 (51.9)	.79
Confidentiality	51 (26.7)	24 (47.1)	27 (52.9)	.67
Quality of information	45 (23.6)	18 (40)	27 (60)	.18
Immediate feedback	41 (21.5)	19 (46.3)	22 (53.7)	.64

[a]STI: sexually transmitted infection.

[b]SRH: sexual reproductive health.

Discussion

Principal Findings

This study explored the use of a USSD-based mobile phone intervention to deliver on-demand adolescent SRH information in an RCT. We studied the effectiveness and impact of a USSD-based mobile phone app on increasing adolescents' knowledge of contraceptives, gender-based stereotypes, STIs, abstinence, and perceived vulnerability. We also evaluated the USSD-based ability of the mobile phone app to help adolescents make informed decisions regarding their SRH. Our results show improved awareness of SRH information and improved knowledge about contraceptives and abstinence. Increased awareness has enabled more adolescents to abstain from sex, improve condom use, and identify STIs. Confidentiality when accessing SRH information was of particular importance to the participants.

Adolescents' needs for information on contraceptives is unmet in most resource-limited settings; adolescents are unable to secure information on available contraceptive options or discover where they can access this information [27]. In our study, adolescents using the app improved their knowledge of contraceptives, with a trend toward statistical significant (*P*=.06). Our findings are promising, and mobile phone apps

could help increase awareness on and knowledge of contraceptives among adolescents. The provision of information on contraceptives to adolescents is complex because of cultural, religious, and political setbacks. Innovative approaches are needed to meet adolescents' information needs. The study outcomes also show the need to make information about contraceptives accessible to adolescents in a culturally and age-appropriate manner [28,29].

When accessing SRH information and services, adolescents want their confidentiality to be respected and upheld. Fear of being *judged* and the possibility of negative attitudes from health care providers can prevent adolescents from accessing these important services [30]. During follow-up visits, of the 62 adolescent participants who had used the mobile app, 51 (82.3%) indicated that confidentiality was one of the most important features of the app. Adolescent users can access any SRH information in a user-friendly manner. Research has shown that adolescents value confidentiality when accessing SRH information and are more willing to seek SRH care and interventions when their confidentiality is assured [31].

mHealth apps have shown great potential for engaging with and increasing SRH information access for adolescents from different age groups and social demographics [32,33]. In one study, text messages improved SRH outcomes by reducing pregnancy rates [34]. The aforementioned studies demonstrate the great potential of mHealth apps in improving and increasing adolescents' knowledge of SRH. Our study findings show that adolescents require high-quality SRH information provided in an easy-to-use, confidential manner with immediate feedback. The USSD technology enables an interactive user-driven mobile app to provide information based on a user's inputs. This USSD technology is low cost, works on both feature phones and smartphones, and can be provided free of charge.

Limitations

During the study, 47 adolescents were unable to use the mobile app, mainly because of a lack of access to mobile phones. This may explain why there appeared to have been a minimal change in adolescent users' knowledge scores. Access to mobile phones in most resource-limited settings is associated with the household economic status. In addition, access to a phone was self-reported. Several adolescents hoped to be provided with a phone by their parents, caregivers, or older siblings. Adolescent participants in the intervention group who were unable to use the app reported that either their parents traveled or the mobile phone they hoped to use stopped functioning. Some studies have opted to provide adolescent participants with mobile phones to ensure that participants in the intervention group accessed the mobile apps. This approach has increased the cost of the study, and other researchers have viewed providing mobile phones as an inducement. In resource-limited settings such as Kenya, access to the internet is limited and web-based apps may not be an option in this setting. However, internet cybercafes are available in many places. Providing internet payment vouchers to adolescents to access the internet and a customized web-based study app could be explored. The results of our study may not be generalizable across Kibra.

Conclusions

Adolescents require accurate and up-to-date SRH information to guide their decision-making and improve health outcomes. As they already use mobile phones in their day-to-day lives, mobile phone apps provide an ideal platform. Considerable promise has been demonstrated by studies using mobile apps to improve adolescents' access to SRH information. Scaled-up research on mHealth apps providing SRH information is required to better evaluate their impact on SRH outcomes.

Acknowledgments

The authors would like to thank James Serembe, Josephine Ocham, Samwel Oninga, and Nolline Oudu for working closely with the adolescents during enrollment and follow-up. In addition, they would like to thank all adolescent participants.

Authors' Contributions

PM, APN, II, JK, RN, and CC contributed to research protocol development, data review, and preparation of this manuscript. BS contributed to data review and paired sample t-test analysis.

Conflicts of Interest

None declared.

Editorial notice: This randomized study was only retrospectively registered. The editor granted an exception of ICMJE rules for prospective registration of randomized trials because the risk of bias appears low and the study was considered formative. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1
Unstructured Supplementary Service Data app content.
[PDF File (Adobe PDF File), 65 KB - mhealth_v10i4e31233_app1.pdf]

Multimedia Appendix 2

Adolescent recruitment script page of the approved consent form in English.
[PDF File (Adobe PDF File), 47 KB - mhealth_v10i4e31233_app2.pdf]

Multimedia Appendix 3
Evaluation of knowledge of sexual reproductive health (SRH) information.
[PDF File (Adobe PDF File), 43 KB - mhealth_v10i4e31233_app3.pdf]

Multimedia Appendix 4
Use and perceptions of the mobile phone app.
[PDF File (Adobe PDF File), 42 KB - mhealth_v10i4e31233_app4.pdf]

Multimedia Appendix 5
CONSORT-EHEALTH (V 1.6.1) checklist.
[PDF File (Adobe PDF File), 936 KB - mhealth_v10i4e31233_app5.pdf]

References

1. Temmerman M, Khosla R, Say L. Sexual and reproductive health and rights: a global development, health, and human rights priority. Lancet 2014 Aug;384(9941):e30-e31. [doi: 10.1016/s0140-6736(14)61190-9]
2. Gonsalves L, Hindin MJ, Bayer A, Carcamo CP, Gichangi P, Habib N, et al. Protocol of an open, three-arm, individually randomized trial assessing the effect of delivering sexual and reproductive health information to young people (aged 13-24) in Kenya and Peru via mobile phones: adolescent/youth reproductive mobile access and delivery initiative for love and life outcomes (ARMADILLO) study stage 2. Reprod Health 2018 Jul 11;15(1):126 [FREE Full text] [doi: 10.1186/s12978-018-0568-6] [Medline: 29996854]
3. Starrs AM, Ezeh AC, Barker G, Basu A, Bertrand JT, Blum R, et al. Accelerate progress—sexual and reproductive health and rights for all: report of the Guttmacher– Lancet Commission. Lancet 2018 Jun;391(10140):2642-2692. [doi: 10.1016/s0140-6736(18)30293-9]
4. WHO Recommendations on Adolescent Sexual and Reproductive Health and Rights. Geneva: World Health Organization; 2018.
5. Platforms for delivering adolescent health actions. In: Child and Adolescent Health and Development. 3rd edition. Washington (DC): The International Bank for Reconstruction and Development / The World Bank; 2017.
6. Lopez C, Ramirez DC, Valenzuela JI, Arguello A, Saenz JP, Trujillo S, et al. Sexual and reproductive health for young adults in Colombia: teleconsultation using mobile devices. JMIR Mhealth Uhealth 2014 Sep 25;2(3):e38 [FREE Full text] [doi: 10.2196/mhealth.2904] [Medline: 25263432]
7. Ybarra ML, Prescott TL, Philips GL, Bull SS, Parsons JT, Mustanski B. Iteratively developing an mHealth HIV prevention program for sexual minority adolescent men. AIDS Behav 2016 Jun;20(6):1157-1172 [FREE Full text] [doi: 10.1007/s10461-015-1146-3] [Medline: 26238038]
8. Mutebi M, Bhatia R, Salako O, Rubagumya F, Grover S, Hammad N. Innovative use of mHealth and clinical technology for oncology clinical trials in Africa. JCO Glob Oncol 2020 Jun;6:948-953 [FREE Full text] [doi: 10.1200/JGO.19.00191] [Medline: 32614724]
9. Giorgio MM, Kantor LM, Levine DS, Arons W. Using chat and text technologies to answer sexual and reproductive health questions: planned parenthood pilot study. J Med Internet Res 2013 Sep 20;15(9):e203 [FREE Full text] [doi: 10.2196/jmir.2619] [Medline: 24055754]
10. Yoost JL, Starcher RW, King-Mallory RA, Hussain N, Hensley CA, Gress TW. The use of telehealth to teach reproductive health to female rural high school students. J Pediatr Adolesc Gynecol 2017 Apr;30(2):193-198. [doi: 10.1016/j.jpag.2016.10.002] [Medline: 27742427]
11. Chen E, Mangone ER. A systematic review of apps using mobile Criteria for Adolescent Pregnancy Prevention (mCAPP). JMIR Mhealth Uhealth 2016 Nov 10;4(4):e122 [FREE Full text] [doi: 10.2196/mhealth.6611] [Medline: 27833070]
12. L'Engle KL, Mangone ER, Parcesepe AM, Agarwal S, Ippoliti NB. Mobile phone interventions for adolescent sexual and reproductive health: a systematic review. Pediatrics 2016 Sep;138(3):e20160884. [doi: 10.1542/peds.2016-0884] [Medline: 27553221]
13. Rokicki S, Fink G. Assessing the reach and effectiveness of mHealth: evidence from a reproductive health program for adolescent girls in Ghana. BMC Public Health 2017 Dec 20;17(1):969 [FREE Full text] [doi: 10.1186/s12889-017-4939-7] [Medline: 29262823]
14. Aicken CR, Fuller SS, Sutcliffe LJ, Estcourt CS, Gkatzidou V, Oakeshott P, et al. Young people's perceptions of smartphone-enabled self-testing and online care for sexually transmitted infections: qualitative interview study. BMC Public Health 2016 Sep 13;16(1):974 [FREE Full text] [doi: 10.1186/s12889-016-3648-y] [Medline: 27624633]

15. Ippoliti NB, L'Engle K. Meet us on the phone: mobile phone programs for adolescent sexual and reproductive health in low-to-middle income countries. Reprod Health 2017 Jan 17;14(1):11 [FREE Full text] [doi: 10.1186/s12978-016-0276-z] [Medline: 28095855]
16. McCarthy OL, Wazwaz O, Osorio Calderon V, Jado I, Saibov S, Stavridis A, et al. Development of an intervention delivered by mobile phone aimed at decreasing unintended pregnancy among young people in three lower middle income countries. BMC Public Health 2018 May 02;18(1):576 [FREE Full text] [doi: 10.1186/s12889-018-5477-7] [Medline: 29716571]
17. Olsen PS, Plourde KF, Lasway C, van Praag E. Insights from a text messaging-based sexual and reproductive health information program in Tanzania (m4RH): retrospective analysis. JMIR Mhealth Uhealth 2018 Nov 01;6(11):e10190 [FREE Full text] [doi: 10.2196/10190] [Medline: 30389651]
18. Macharia P, Pérez-Navarro A, Inwani I, Nduati R, Carrion C. Developing a USSD-based mobile phone app to provide adolescents with sexual reproductive health information: a human-centered design approach. JMIR 2021 (forthcoming) [FREE Full text] [doi: 10.2196/preprints.29089]
19. Macharia P, Pérez-Navarro A, Inwani I, Nduati R, Carrion C. An exploratory study of current sources of adolescent sexual and reproductive health information in Kenya and their limitations: are mobile phone technologies the answer? Int J Sexual Health 2021 May 16;33(3):357-370. [doi: 10.1080/19317611.2021.1918311]
20. The health belief model. In: The Wiley Encyclopedia of Health Psychology. Hoboken, New Jersey, United States: Wiley; 2020.
21. Phillips K. Review of the AVERT, a global information and education on HIV and AIDS resource. J Consumer Health Internet 2019 Sep 13;23(3):290-298. [doi: 10.1080/15398285.2019.1648162]
22. Kenya adolescent reproductive health and development policy: implementation assessment report. UNESCO. URL: https://healtheducationresources.unesco.org/library/documents/kenya-adolescent-reproductive-health-and-development-policy-implementation [accessed 2022-03-13]
23. Clark L, Fairhurst C, Torgerson DJ. Allocation concealment in randomised controlled trials: are we getting better? BMJ 2016 Nov 17;355:i5663. [doi: 10.1136/bmj.i5663] [Medline: 27856428]
24. Create a blocked randomisation list. Sealed Envelope. URL: https://www.sealedenvelope.com/simple-randomiser/v1/lists [accessed 2022-03-13]
25. Percent of adolescents who have “positive” attitudes toward key sexual and reproductive health issues. US AID. URL: https://www.data4impactproject.org/prh/womens-health/adolescent-and-youth-sexual-and-reproductive-health/percent-of-adolescents-who-have-positive-attitudes-toward-key-sexual-and-reproductive-health-issues/ [accessed 2022-03-13]
26. Illustrative questionnaire for interview-surveys with young people. World Health Organization. URL: https://www.scirp.org/(S(czeh2tfqyw2orz553k1w0r45))/reference/ReferencesPapers.aspx?ReferenceID=1950110 [accessed 2022-03-13]
27. Sidibé S, Delamou A, Camara BS, Dioubaté N, Manet H, El Ayadi AM, et al. Trends in contraceptive use, unmet need and associated factors of modern contraceptive use among urban adolescents and young women in Guinea. BMC Public Health 2020 Dec 01;20(1):1840 [FREE Full text] [doi: 10.1186/s12889-020-09957-y] [Medline: 33261605]
28. Li Z, Patton G, Sabet F, Zhou Z, Subramanian SV, Lu C. Contraceptive use in adolescent girls and adult women in low- and middle-income countries. JAMA Netw Open 2020 Feb 05;3(2):e1921437 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.21437] [Medline: 32074290]
29. Rizvi F, Williams J, Bowe S, Hoban E. Factors influencing unmet need for contraception amongst adolescent girls and women in Cambodia. PeerJ 2020;8:e10065 [FREE Full text] [doi: 10.7717/peerj.10065] [Medline: 33083131]
30. Leichliter JS, Copen C, Dittus PJ. Confidentiality issues and use of sexually transmitted disease services among sexually experienced persons aged 15-25 years - United States, 2013-2015. MMWR Morb Mortal Wkly Rep 2017 Mar 10;66(9):237-241 [FREE Full text] [doi: 10.15585/mmwr.mm6609a1] [Medline: 28278143]
31. Pampati S, Liddon N, Dittus PJ, Adkins SH, Steiner RJ. Confidentiality matters but how do we improve implementation in adolescent sexual and reproductive health care? J Adolesc Health 2019 Sep;65(3):315-322 [FREE Full text] [doi: 10.1016/j.jadohealth.2019.03.021] [Medline: 31227388]
32. Rokicki S, Fink G. Assessing the reach and effectiveness of mHealth: evidence from a reproductive health program for adolescent girls in Ghana. BMC Public Health 2017 Dec 20;17(1). [doi: 10.1186/s12889-017-4939-7]
33. Feroz AS, Ali NA, Khoja A, Asad A, Saleem S. Using mobile phones to improve young people sexual and reproductive health in low and middle-income countries: a systematic review to identify barriers, facilitators, and range of mHealth solutions. Reprod Health 2021 Jan 16;18(1):9 [FREE Full text] [doi: 10.1186/s12978-020-01059-7] [Medline: 33453723]
34. Chernick LS. Improving adolescent sexual and reproductive health: can mobile health interventions affect behavior? Pediatrics 2021 Mar 10;147(3):e2020029801 [FREE Full text] [doi: 10.1542/peds.2020-029801] [Medline: 33568492]

Abbreviations

mHealth: mobile health
RCT: randomized controlled trial
SRH: sexual reproductive health
STI: sexually transmitted infection

USSD: Unstructured Supplementary Service Data
WHO: World Health Organization

Edited by L Buis, A Mavragani; submitted 14.06.21; peer-reviewed by D Levine, J Cyriac; comments to author 05.08.21; revised version received 07.09.21; accepted 20.02.22; published 15.04.22.

Please cite as:
Macharia P, Pérez-Navarro A, Sambai B, Inwani I, Kinuthia J, Nduati R, Carrion C
An Unstructured Supplementary Service Data–Based mHealth App Providing On-Demand Sexual Reproductive Health Information for Adolescents in Kibra, Kenya: Randomized Controlled Trial
JMIR Mhealth Uhealth 2022;10(4):e31233
URL: https://mhealth.jmir.org/2022/4/e31233
doi:10.2196/31233
PMID:35436230

Original Paper

Predicting Psychotic Relapse in Schizophrenia With Mobile Sensor Data: Routine Cluster Analysis

Joanne Zhou[1], BA; Bishal Lamichhane[2], MS; Dror Ben-Zeev[3], PhD; Andrew Campbell[4], PhD; Akane Sano[2], PhD

[1]Department of Statistics, Rice University, Houston, TX, United States

[2]Department of Electrical and Computer Engineering, Rice University, Houston, TX, United States

[3]Behavioral Research in Technology and Engineering Center, Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, United States

[4]Department of Computer Science, Dartmouth College, Hanover, NH, United States

Corresponding Author:
Joanne Zhou, BA
Department of Statistics
Rice University
6100 Main Street
Houston, TX, 77005
United States
Phone: 1 5715194641
Email: joanneyz98@gmail.com

Abstract

Background: Behavioral representations obtained from mobile sensing data can be helpful for the prediction of an oncoming psychotic relapse in patients with schizophrenia and the delivery of timely interventions to mitigate such relapse.

Objective: In this study, we aim to develop clustering models to obtain behavioral representations from continuous multimodal mobile sensing data for relapse prediction tasks. The identified clusters can represent different routine behavioral trends related to daily living of patients and atypical behavioral trends associated with impending relapse.

Methods: We used the mobile sensing data obtained from the CrossCheck project for our analysis. Continuous data from six different mobile sensing-based modalities (ambient light, sound, conversation, acceleration, etc) obtained from 63 patients with schizophrenia, each monitored for up to a year, were used for the clustering models and relapse prediction evaluation. Two clustering models, Gaussian mixture model (GMM) and partition around medoids (PAM), were used to obtain behavioral representations from the mobile sensing data. These models have different notions of similarity between behaviors as represented by the mobile sensing data, and thus, provide different behavioral characterizations. The features obtained from the clustering models were used to train and evaluate a personalized relapse prediction model using balanced random forest. The personalization was performed by identifying optimal features for a given patient based on a personalization subset consisting of other patients of similar age.

Results: The clusters identified using the GMM and PAM models were found to represent different behavioral patterns (such as clusters representing sedentary days, active days but with low communication, etc). Although GMM-based models better characterized routine behaviors by discovering dense clusters with low cluster spread, some other identified clusters had a larger cluster spread, likely indicating heterogeneous behavioral characterizations. On the other hand, PAM model-based clusters had lower variability of cluster spread, indicating more homogeneous behavioral characterization in the obtained clusters. Significant changes near the relapse periods were observed in the obtained behavioral representation features from the clustering models. The clustering model-based features, together with other features characterizing the mobile sensing data, resulted in an F2 score of 0.23 for the relapse prediction task in a leave-one-patient-out evaluation setting. The obtained F2 score was significantly higher than that of a random classification baseline with an average F2 score of 0.042.

Conclusions: Mobile sensing can capture behavioral trends using different sensing modalities. Clustering of the daily mobile sensing data may help discover routine and atypical behavioral trends. In this study, we used GMM-based and PAM-based cluster models to obtain behavioral trends in patients with schizophrenia. The features derived from the cluster models were found to be predictive for detecting an oncoming psychotic relapse. Such relapse prediction models can be helpful in enabling timely interventions.

(JMIR Mhealth Uhealth 2022;10(4):e31006) doi:10.2196/31006

KEYWORDS

schizophrenia; psychotic relapse; machine learning; clustering; mobile phone; routine; Gaussian mixture models; partition around medoids; dynamic time warping; balanced random forest

Introduction

Background

Schizophrenia is the most common psychotic disorder, affecting up to 20 million people worldwide [1] and accounting for more than 13.4 million years of life lived with a disability [2]. It can be caused by a combination of genetic, environmental, and psychosocial factors. Patients with schizophrenia experience a range of positive symptoms (hallucinations, delusions, etc), negative symptoms (anhedonia, social withdrawal, etc), and cognitive dysfunctions (lack of attention, working memory, executive function, etc) [3,4]. The disorder is highly disabling and often has consequences such as impairment of education, employment, and social contact [4]. Adults with schizophrenia also have an increased risk of premature mortality than the general population [5]. Therefore, proper treatment and management of schizophrenia are important to limit the negative impact of the disorder on the individual's life.

Schizophrenia is usually treated with a combination of antipsychotic medications and psychosocial treatments. However, patients undergoing treatment can still experience psychotic or symptomatic relapse, an acute exacerbation of schizophrenia symptoms [6]. A previous study found that the cumulative first and second relapse rates were 81.9% and 78%, respectively, within 5 years of recovery from the first episode of schizophrenia and schizoaffective disorder [7]. The risk of relapse is found to be significantly higher after treatment reduction or discontinuation [6]. Relapse poses severe health risks for the individuals and can jeopardize their treatment progression and daily functioning. Each relapse episode is associated with a risk of self-harm and harm to others [8].

To monitor a patient's health status and recovery, routine clinic visits for continual assessment are recommended. Clinical interviews and questionnaire tools were used during the visit to assess current health symptoms and provide timely intervention to prevent relapses [9]. However, relapses may occur between the visits, during which a patient's health status is not assessed. In addition, patients may have limited insight during a psychotic relapse and may struggle to report it to the treatment team or a significant other. Therefore, improving treatment adherence and preventing relapses have become a focus of schizophrenia management. Toward the effort of relapse prevention, there has been significant interest in mobile sensing-based behavioral monitoring models for automatic relapse risk prediction.

Previous Studies

Smartphone apps and wearable devices have been used in several previous studies to collect passive sensing data and track daily behaviors, which can then be used to model the relationship between behaviors and mental well-being. For example, in the StudentLife study, an Android sensing app collected passive sensing data from 48 college students, and the inferred behavioral features from the collected data were found to be correlated with academic performance and self-reported mental health conditions [10]. In a study on depression severity, the mobile sensing-based features, such as daily behavioral rhythms, variance of patient's location, and phone use, were found to be related to depressive symptom severity [11]. The use of mobile sensing to collect long-term monitoring data has also been demonstrated to be feasible and acceptable for patients with schizophrenia disorders [12-15]. Surveys have found that people with schizophrenia commonly access digital devices for communication and support related to the disorder, which again shows the applicability of using mobile sensing as a platform to monitor schizophrenia symptoms [16].

Mobile sensing data have been used to model behaviors and predict psychotic relapses in patients with schizophrenia. If an oncoming relapse can be detected with high accuracy, timely medical interventions can be provided to mitigate the associated risks. Researchers have found anomalies in daily behavior assessed from mobile sensing before relapses and developed relapse prediction models with promising accuracy [17-19]. In a pilot study, the Beiwe app collected mobile sensing data from 15 patients with schizophrenia for 3 months, during which 5 patients experienced relapses [17]. The researchers found that the rate of anomalies in mobility and social behavior increased significantly closer to relapses. In the CrossCheck project, a mobile sensing app was developed to collect self-reporting ecological momentary assessment and continuous passive sensing data from 75 outpatients with schizophrenia [20]. On the basis of this data set, Wang et al [18] compared different machine learning models for relapse prediction, with several feature extraction windows, and identified the best classifier and prediction settings for detecting an oncoming relapse. The best performance was obtained using a support vector machine (with radial basis function kernel) model and a feature extraction window of 30 days, leading to an F1 score of 0.27 on the relapse prediction task. Similarly, the Adler et al [21] used an anomaly detection framework based on an encoder-decoder reconstruction loss to predict psychotic relapse in schizophrenia.

Concerning current mental health status, the extent to which an individual adheres to work, sleep, social, or mobility routine (ie, a regular behavioral pattern) largely impacts their mental well-being and symptom severity of mental disorders [11,22,23]. Behavioral stability features that measure the adherence to routines have been proposed as relapse predictors in some of the previous studies. Features computed in our previous study measured behavioral stability by calculating the temporal evolution of daily templates of features derived from mobile sensing data (daily templates are time series obtained with representative feature values at regular time intervals in a given day; eg, time series of hourly feature values) [19]. Tseng et al [24] also showed the effectiveness of using behavioral

rhythm-based features to predict different symptom severities. Stability features such as deviation of daily templates were found to be significant predictors of schizophrenia symptoms, such as depression. He-Yueya et al [25] also proposed a stability metric for behaviors with a fine temporal resolution by calculating the distance between 2 cumulative sum functions describing behaviors in a certain minute of the day. The computed stability index had similar predictive power as the state-of-the-art behavioral features (mean and SD of each behavior) in the study by Wang et al [26], while being complementary. In all the previous studies using behavioral stability to model relapse prediction, the measured stability was limited to the behaviors observed within a short feature extraction window (eg, few weeks only). An individual's routine behaviors were not fully represented owing to the short time window considerations. A summary of behavioral patterns can rather be obtained when large time windows are considered.

In this study, instead of measuring behavioral patterns using the variance of day-to-day behaviors, we identify the overall cluster of behaviors for an individual using multimodal mobile phone data and unsupervised machine learning and derive features based on the distance of behaviors observed in a day compared with the individual's most representative routines. The identified behavioral clusters for an individual could represent their weekday routine, weekend routine, low-phone-use routine (no sensor reading), and so on. The identified clusters provide a representation of the long-term behavioral trends across the patients, which are not directly captured by short-term behavioral rhythm features, as used in previous studies. Furthermore, clusters obtained from the mobile sensing data represent quantized behaviors, and features derived from these clusters are robust to the insignificant variations in behavior compared with the short-term behavioral rhythm change features. Typical behavioral routines for individuals can be found via the clustering analysis of their daily behaviors. Previously, clustering has been applied for identifying mobility patterns using GPS sensing data and evaluating anomalies accordingly [21,26]. However, to the best of our knowledge, clustering analysis has not been performed for characterizing the overall behavioral patterns of patients with schizophrenia, using multimodal mobile sensing data toward relapse prediction tasks.

Goal of This Study

In this study, we aim to (1) develop a method to characterize patients' daily behaviors using multimodal smartphone sensor data, (2) understand the relationship between behavioral patterns and psychotic relapse events in schizophrenia, and (3) evaluate the predictive power of the identified behavioral pattern-based features for relapse prediction. We propose multivariate time series clustering of daily templates obtained from mobile sensing data to obtain behavioral patterns. Then, the features derived from clustering are used in the relapse prediction task. The paper is organized as follows. In the *Methods* section, we describe the method used to cluster multidimensional daily templates from mobile sensing data, model selection approach for clustering, and feature extraction and relapse prediction modeling. In the *Results* section, we present the results obtained from the clustering models, association of the obtained clustering-based behavioral features with relapses, and evaluation of the developed relapse prediction model. The obtained results are discussed and future directions are outlined in the *Discussion* section.

Methods

Ethics Approval

This study was approved by the ethical review committee of Dartmouth College (#24356) and the institutional review board of North Shore-Long Island Jewish Health System (#14-100B) [20].

Data Preparation

The data used in this study were obtained from the CrossCheck project (clinical trial registration: ClinicalTrials.gov, NCT01952041 [27]), which was conducted at the Zucker Hillside Hospital in New York City, New York [20,24,26,28,29]. Informed consents were obtained from the participants. The inclusion criteria for the participants are described in the study by Ben-Zeev et al [20]. The CrossCheck app collected mobile sensing data from 75 outpatients with schizophrenia with a data collection period of >12 months per patient. Of the 75 patients, 63 (84%) patients completed the data collection (n=27, 43% men and n=36, 57% women; average age 37.2, SD 13.7 years; range 18-65 years), and a total of 27 relapse events occurred in 32% (20/63) of the patients during the monitoring period. Some patients had multiple incidences of relapses, but as the monitoring period was long, each of the incidences was treated as a unique event if separated by a month. A relapse incident was defined as one that has occurred under one or more of the following seven criteria: psychiatric hospitalization, increased frequency or intensity of services, increased medications or dosages or >25% changes in Brief Psychiatric Rating Scale scores, suicidal ideation, homicidal ideation, self-injury, and violent behavior resulting in harm to self or others [18]. A total of 6 mobile sensing modalities including physical activity, sociability, and ambient environmental readings were obtained using the app. Different features were extracted from these mobile sensing modalities, as presented by Tseng et al [24]. From these features, a total of 21 passive sensing features were selected for our proposed clustering-based behavioral characterization: acceleration, distance traveled, sleep duration, ambient sound, ambient light, conversation duration, phone unlock duration, and different types of call log, SMS text message log, and app use. All the features were transformed to an hourly resolution by averaging the observations within 1 hour. For features that were obtained with lower resolution (eg, every few hours), for example, distance traveled from morning to noon, the feature values were split to each hour spanned by the time represented by these feature values. With hourly resolution for each of the 21 features considered and these hourly feature values considered as separate feature space, the resulting data set had a dimension of 504 (21×24). Observational data for a total of 18,436 days were collected for all the patients. Per-patient feature normalization (min-max normalization between 0 to 1) was performed to adjust for differences between patients. From the normalized data set, principal component analysis on the full data set (with data from all the patients) was

performed for dimensionality reduction. The first 200 principal components were retained, which explained 96.9% of the total variance.

Clustering Models

We evaluated two different clustering methods: Gaussian mixture model (GMM) and partition around medoids (PAM), to cluster the features from the mobile sensing data and obtain behavioral representations. The 2 clustering models differ in how the similarity between different points are assessed, representing different ways in which behaviors across days can be compared with each other, and therefore, produce different cluster outputs.

GMM Clustering

Model Introduction

The GMM is a probabilistic model that assumes data are generated from a finite set of Gaussian distributions. Gaussian mixture probability density is the weighted sum of k component Gaussian densities [30]. The GMM can address correlation between attributes by selecting the optimal covariance matrix for each cluster and has been used in previous behavioral clustering problems [31]. Moreover, it can derive the probability of each sample in its assigned Gaussian distribution. In this study, we used the GMM implementation from the scikit-learn package in Python to obtain a clustering model for the mobile sensing data [32]. The parameters of the GMM were obtained using the expectation-maximization algorithm [33]. We selected the number of clusters and covariance matrix type based on Akaike information criterion (AIC) score and Bayesian information criterion (BIC) score of all the candidate models (see more details in Multimedia Appendix 1 [34,35]).

Model Output

Three output variables for each of the data points (observations), offering GMM-based clustering features for the data points, were generated based on the developed GMM: cluster label, assigned cluster likelihood score, and weighted average likelihood score.

Cluster label is represented by integers from 1 to k *(*k is the number of clusters selected in the GMM*)*. Cluster likelihood scores derived from the model measure how *irregular* each day (represented by a data point) is by calculating its deviation from the Gaussian mixtures. If we consider the center of each Gaussian as a typical routine, the farther out a point is in this Gaussian space, the higher the chances that the point represents an anomalous day or behavior.

The likelihood of a data point in a multivariate Gaussian distribution can be computed by calculating the probability of observing a point farther than the given point. In other words, the cumulative distribution function is evaluated at a given data point, which can be obtained using Mahalanobis distance metric. Note that the squared Mahalanobis distance from a point to the center of a Gaussian distribution has been proven to follow a chi-squared distribution with p df, where p is the number of variables [36]. Therefore, the likelihood of a point in the Gaussian distribution is equivalent to the cumulative probability of observing a value larger than the given Mahalanobis distance in a chi-squared distribution with p df.

The assigned cluster likelihood score of the data point was obtained as the probability of each point to its assigned cluster. The weighted average likelihood score was computed as the weighted (with the cluster's corresponding weights) sum of the probability of a given point belonging to each of the Gaussian classes. Intuitively, the assigned cluster likelihood score measures how similar a day is to its closest routine. The weighted average likelihood score measures how similar a day is to all routines. As the weighted average likelihood score accounts for cluster weights, a point that is closer to a more populous cluster will be considered less anomalous. A 2D illustration of the likelihood scores is provided in Figure S1 in Multimedia Appendix 1.

PAM With Dynamic Time Warping Clustering

Model Introduction

GMM measure the similarity between observations (data points) using point-wise alignment of different features in the observation. However, the dissimilarity between 2 observations could be overestimated owing to an outlier (eg, because of faulty sensor measurements) or when there is a small time-shift or speed difference between observations. For example, 2 daily templates with a similar pattern but a shift of 1 hour would be expected to represent similar behavioral representations, but these templates would likely be considered dissimilar from the GMM. To allow flexible similarity assessments, we used dynamic time warping (DTW) to determine the optimal alignment of indices of the 2 time series that minimizes the distance between the time series [37]. The DTW distance can be paired with a distance-based clustering method such as a PAM clustering model [38]. The PAM model searches for k representative objects (medoids) from the data and creates clusters so that the total dissimilarity of points within clusters is minimized. We compared the number of clusters k based on the sum of the squared DTW distance of every data point with its cluster medoid and the elbow method (see more details in Multimedia Appendix 1).

Model Output

From the fitted PAM model, similar to the procedure after GMM fit, we generated three output features characterizing each data point: cluster label, assigned cluster distance score, and weighted average distance score. Similar to the GMM-based likelihood score computation, the cluster distance scores evaluate how dissimilar each object is from a representative data point or from all representative data points. The assigned cluster distance score is the DTW distance of each data point (representing a daily template) to its cluster medoid. A lower value indicates that a day conforms better to its most similar routine. Weighted average distance score was obtained by summing the DTW distance to all medoids scaled by the corresponding cluster sizes. A lower value indicates that a day conforms better to all possible routines. DTW distance from the previous day's daily template was also calculated as a potential relapse predictor.

Analyzing Cluster Results

After obtaining the output variables from the cluster models, we evaluated whether there were significant changes in any of these cluster output variables closer to relapse events. To quantify this change, we first defined different key periods to focus before a relapse. Similar to a previous study, we defined NRx as x days near relapse period (before the relapse event) and pre-NRx as all days before relapses that were not in NRx (healthy period) [21]. We evaluated the cluster outputs for NR7, NR14, NR20, and NR30 periods to test different window sizes. Cliff δ was computed to estimate the size of the change in the likelihood scores (GMM output) and distance scores (PAM model output) between the NRx and pre-NRx periods for each patient separately [39]. Cliff δ was chosen because of the nonnormality and variance heterogeneity of our data, for which the Cliff δ is a suitable metric. It is calculated as follows:

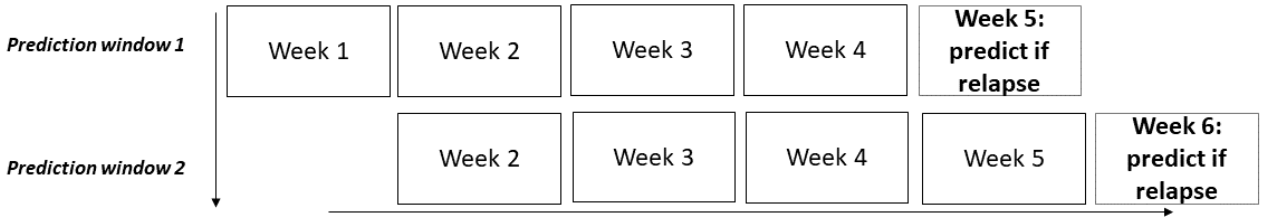
In the formula, $\#(x_1 > x_2)$ counts the number of values in group 1 (NRx period) that is larger than a value in group 2 (pre-NRx period) for all value pairs, and n_1 and n_2 are the sample sizes. This effect size ranges from −1 to 1, where 1 indicates that all values in the NRx period are larger than all values in the pre-NRx period, and −1 indicates vice-versa. As proposed by Romano et al [40], the effect can be considered to be nonnegligible if the absolute value is >0.147. Cliff δ is suitable to compare continuous variable output such as likelihood scores and distance scores.

Relapse Prediction

Relapse Prediction Approach

We framed relapse prediction as a binary classification problem similar to the earlier studies [19,41]. On the basis of the mobile sensing features derived from a feature extraction window (current and immediately past observations from a patient), we predicted whether the patient is likely to experience a relapse in an oncoming period (prediction window). Similar to previous studies [19,41], we used a 4-week period as the feature extraction window and a 1-week period as the prediction window (Figure 1). Thus, mobile sensing observations from the past 4-week period were used in the relapse prediction model to predict whether there will be a relapse in the next week.

Figure 1. Sequential relapse prediction approach used in this study. Features are extracted from a period of 4 weeks to predict if a relapse might occur in the coming week.



Features

Overview

Mobile sensing data contain features that characterize behavioral patterns in the relapse prediction model. In this study, we evaluated the contribution of the clustering features derived from the GMM and PAM models for the psychotic relapse prediction task. We briefly describe the baseline features (based on an earlier study [19]) and clustering-based features that are added to the relapse prediction model.

Baseline Features

These consist of all the features as used in the study by Lamichhane et al [19] along with distance-based and duration-based mobility features and screen use-based features. The CrossCheck data set contains information about when the screen of the patient's smartphone is active. A single screen-use modality that represents the time spent using the phone (phone screen was active) was derived. Using this modality, the mean and SD of daily averages in a given feature extraction window were computed as features for the relapse prediction model. Similarly, for mobility-based features, we computed four different mobility modalities: distance traveled from home (home information obtained based on the clustering of the GPS locations), total movement, average time spent in a location, and total time spent at home. Then, for each mobility-based modality, we computed the mean and SD of the daily averages as features characterizing a feature extraction window.

Clustering-Based Features

We extended the baseline feature set with our proposed clustering-based features for the relapse prediction task. These features are listed in Table 1.

Table 1. Features used in relapse prediction models. Baseline features are derived from a previous study [19]. We evaluated if the clustering-based features could improve relapse prediction by complementing the daily behavioral rhythm change-based features represented in the baseline features.

Feature set and modalities	Features
Baseline features	
Accelerometer magnitude; ambient light; distance traveled; call duration; sound level; and conversation duration	Mean daily template features (mean, SD, max, range, skewness, and kurtosis), SD template features (mean), absolute difference between mean and maximum template (max), distance between normalized mean templates, weighted distance between normalized mean templates, distance between normalized maximum template and mean template, and daily averages (mean and SD)
10-item EMAs[a]	Mean and SD of EMA items in feature extraction window
Screen use, distance-based mobility features: distance from home and total movement, and duration-based mobility features: time spent at a location and time spent at home	Mean and SD of daily averages in feature extraction window
Clustering features	
GMM[b] features	Mean and SD of GMM label and GMM likelihood scores, number of cluster transitions, and number of cluster states
PAM[c] features	Mean and SD of PAM label, PAM distance scores, and DTW[d] difference from the previous day; number of cluster transitions; and number of cluster states
Demographic features	Age and education years

[a]EMA: ecological momentary assessment.

[b]GMM: Gaussian mixture model.

[c]PAM: partition around medoids.

[d]DTW: dynamic time warping.

Classifier

For our relapse prediction pipeline, we used a balanced random forest (BRF) classifier with a low overall model complexity (using 11 decision trees). As a classifier, BRF is suitable for learning from an imbalanced data set, as is the case in our relapse prediction task and provides meaningful prediction probabilities in different decision fusion schemes (eg, in situations where only a limited number of sensor modalities are available for a patient). The number of decision trees to be used was heuristically chosen to limit the model size (lower number of trees), while still having a number of trees to maintain diversity for the generalizability of the ensemble model. We used the BRF implemented in the imbalanced-learn library in Python [42], allowing the default unrestricted depth of trees and squared root of number of features considered for best split in the trees. Similar to the approach used by Lamichhane et al [19], features were quantized into discrete bins before being provided as input to the classifier. The number of bins was set as a hyperparameter, and for the set number of equal-width bins, the count of feature values in each bin was retained as the processed feature values. The approach of feature quantization was found to be helpful in relapse prediction, probably by ignoring small insignificant changes while retaining large feature variations representing significant behavioral deviations. We used leave-one-patient-out cross-validation for the evaluation of the model. The number of bins to be used was a hyperparameter for the classification model and was set with cross-validation within the training set (nested cross-validation). The number of bins for feature quantization considered in hyperparameter tuning were 2, 3, 4, 5, 10, and 15, and the tuning procedure is further described in Multimedia Appendix 1.

Relapse Labels

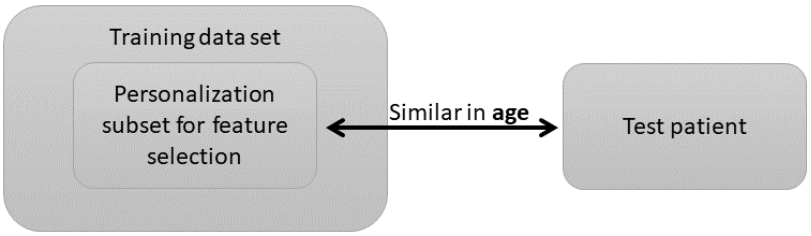
For our relapse prediction pipeline, as the relapse dates are not a fixed discrete event, a hard label and earlier indications of an oncoming relapse are also valuable, we considered the entire month preceding the date of indicated relapse as a relapse period for classification. Thus, any prediction of relapse within a 4-week period before the relapse was considered as a useful output from the prediction model, as has also been used in previous study on relapse prediction [21]. A relapse prediction generated up to a month before the relapse would be observable and potentially actionable for interventions, as behavioral changes associated with relapse could manifest up to a month preceding a relapse [18].

Personalization

Human behavior and behavioral change manifestations of relapse can be person-dependent. Lamichhane et al [19] proposed a method for personalizing a relapse prediction model based on feature selection adapted to a particular test patient. The adaptation occurs using a personalization subset. This is illustrated in Figure 2. For a test patient, within the leave-one-patient-out cross-validation approach, the data from patients closest in age to the given test patient comprised the personalization subset. We included age-based personalization as a first step toward personalized relapse prediction, as behavioral tendencies could be dependent on age, among other factors. Age has been reported to be a significant factor in univariate regression modeling of relapse behaviors in patients with schizophrenia [43], and age dependence of psychosocial functions, substance use behaviors, psychotic symptoms, hospitalization risks, and so on, have been reported in the

context of psychotic relapse in patients with schizophrenia [44]. We evaluated the gains from age-based personalization compared with a nonpersonalized model to empirically establish if age-based personalization could be helpful in behavioral modeling and relapse prediction. As relapse incidents are rare, all the relapse incidents in the training data set were included as a part of the personalization subset. For training a classifier toward the test patient, the optimal features were selected using the personalization subset. We used this approach for training our relapse prediction model, and used the correlation between features and target labels as the feature selection criteria. The number of features to be selected was set as a hyperparameter in our classifier, and this dictated the threshold for correlation value used for feature selection. For example, if the number of features to be selected is 5, the threshold for correlation coefficient (absolute value) is selected such that top-5 features with the highest correlation with the labels are retained. The number of features to be used was selected from 3, 5, 10, and 15, and the features and size of the personalization subset was selected from 50, 75, 100, 125, 150, 200, and 300 for the hyperparameter tuning (further described in Multimedia Appendix 1).

Figure 2. Personalization approach for the relapse prediction model [19]. A personalization subset, consisting of data from patients who are closest in age to the test patients, is used to identify the best feature sets, using which a (personalized) relapse prediction model can be trained.



Evaluation Metric

We evaluated relapse prediction performance to assess the contributions from clustering-based features. Any improvement in the relapse prediction performance when clustering-based features are added to the baseline features would establish the value of clustering-based features to represent behavioral trends and detect anomalies relevant for relapse prediction. Similar to the study by Lamichhane et al [19], we used F2 score for model evaluation to slightly prioritize recall over precision (detecting a relapse is slightly prioritized over generating a false positive). F2 score is given as follows:

We also report the obtained precision and recall scores together with the F2 scores.

Results

Clustering Results

We trained GMM and PAM models to obtain cluster centers and identify different behavioral routine representations. The model selection procedure is explained in Multimedia Appendix 1, and model comparison metrics for GMM and PAM are plotted in Figures S2 and S3 in Multimedia Appendix 1. For GMM, after evaluating AIC and BIC scores, model selection was narrowed to 8 to 14 clusters with full covariance matrix. Among the models with equally good AIC and BIC scores, the models with 9 and 13 clusters achieved the best model stability and least overlap between Gaussian classes. The final model selection was 9 clusters because lower number of clusters allows for higher interpretability. The number of clusters for the PAM model was also selected to be 9 based on the distance dissimilarity metric and elbow method.

See Figures S4 and S5 in Multimedia Appendix 1 for the output from the GMM and PAM models, including cluster size, average likelihood scores (for GMM), distance scores (for PAM; Figure S4 in Multimedia Appendix 1), and kernel density plots that illustrate the distributions of likelihood scores and distance scores (Figure S5 in Multimedia Appendix 1).

To evaluate how well the days in each cluster conform to a routine—the one represented by the cluster center—we measured the spread of each cluster using the trace of the covariance matrix of all cluster samples. Results are illustrated in Figure 3. Clusters with small covariance trace had low within-cluster variability. The GMM cluster model resulted in a more extreme distribution of cluster spread (higher range of covariance trace) because it allowed the clusters to overlap (despite our model selection approach to limit overlaps), whereas the PAM model created partitions in the data.

By averaging all daily templates (data points) in every cluster, it was possible to observe the cluster profiles. For example, Figure 4 illustrates the average daily templates of two example signal modalities: acceleration and conversation. The GMM performed better in stratifying daily templates based on the overall level of activity in these signal modalities. The PAM model had high variance in each cluster because it allows for a more lenient dissimilarity measurement. Although the daily templates in each cluster have different levels of signal activity, they generally follow the same pattern as a normal circadian rhythm; for example, the conversation signal activity peaking during the day and being at minimum during the night.

Table 2 summarizes the average profile for each cluster, ordered from the most common to the least common cluster.

Figure 3. Trace of the sample covariance matrix for each cluster obtained with Gaussian mixture model (GMM) and partition around medoids (PAM) clustering approach. A low covariance matrix trace indicates more homogeneous clusters, that is, clusters with lower within-cluster variability.

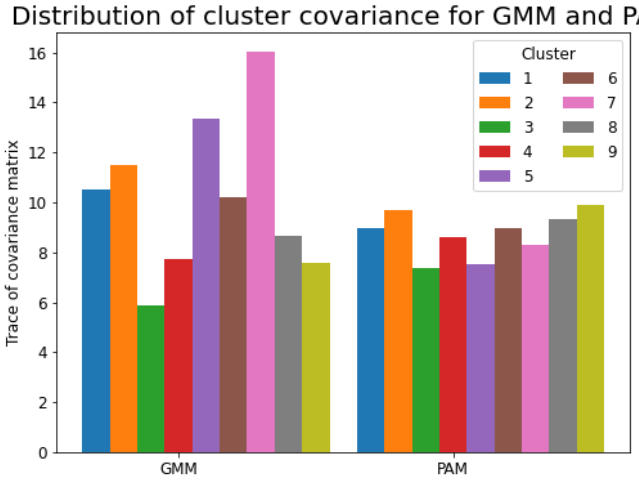


Figure 4. Average daily templates of two signal modalities acceleration (top) and conversation time (bottom) in the clusters obtained from the Gaussian mixture model (GMM) and partition around medoids (PAM) models. Different clusters capture different behavioral patterns.

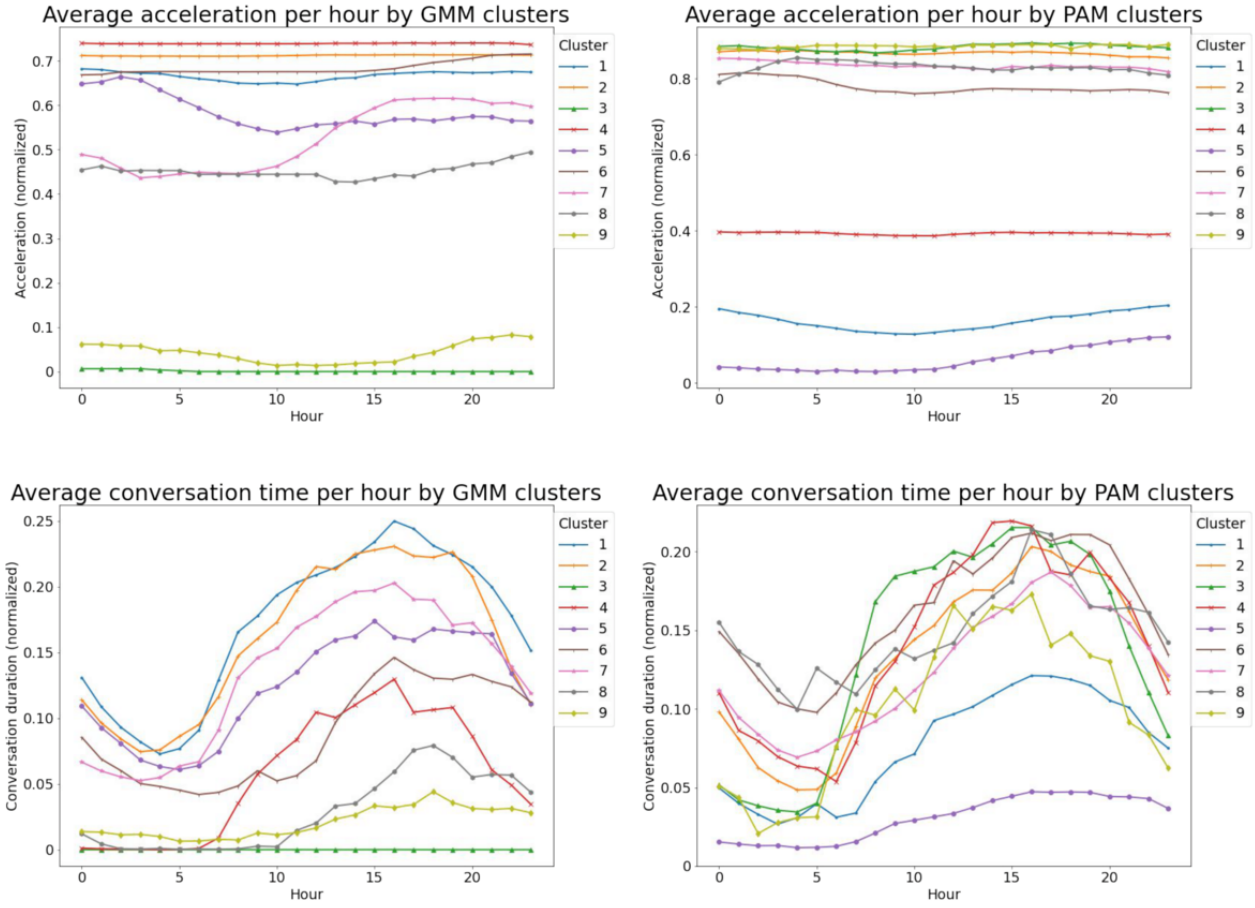


Table 2. All cluster profiles obtained from the GMM[a] and PAM[b] models in descending cluster size. Different clusters are associated with peculiar behaviors specific to that cluster as it can be observed from the typical profile of signal modality in that cluster.

Cluster size rank	GMM cluster profile	GMM cluster size (days)	PAM cluster profile	PAM cluster size (days)
1	No app use, high conversation and SMS text messaging, and other attributes are approximately average	5217	Low acceleration, conversation, volume, and sleep duration and very low variability in sleep and volume templates	3318
2	Highest app use and phone calls; high acceleration, conversation, SMS text messaging, distance moved, and volume; early wake up (at approximately 7 AM); and no sleep during the day	3993	High volume and SMS text messaging and constantly low sleep template	3300
3	Almost all sensor readings near 0	2580	Conversation and volume sharply increase after 6 AM, highest volume, low phone use before 7 AM, wake up at approximately 7 AM, and sleep at approximately 9 PM	2728
4	Highest acceleration, low phone calls, early wake up (at approximately 7 AM), and no sleep during the day	1883	High app use, SMS text messaging, and distance moved around midnight and below average acceleration	2699
5	High acceleration after midnight, high phone calls and SMS text messaging, high overall volume even at night, and late sleep and wake up	1484	Lowest acceleration (close to 0) and app use and constantly high screen time and sleep duration	2378
6	Below average volume and distance, wake up after 11 AM, and sleep during the day	1298	High phone calls and SMS text messaging, screen time sharply increases after 6 AM, wake up at approximately 9 AM, sleep at approximately 11 PM, and awake during the day	1686
7	Activity level and phone use are high during the day, inactive at night, short sleep duration, high number of phone calls, and acceleration increases after 3 PM	1046	Below average screen time and long sleep time (wake up around noon)	1405
8	No app use; low conversation, SMS text messaging, and volume; and long sleep even during the day	523	Low phone call, SMS text messaging, and screen time; high volume at night; and constant long sleep (wake up in the afternoon)	752
9	Accelerometer readings close to 0; low app use, conversation, and volume; phone screen is constantly on; and long sleep duration even during the day	412	High app use and distance moved during noon; templates in this cluster have high dissimilarities	170

[a]GMM: Gaussian mixture model.

[b]PAM: partition around medoids.

Association With Relapses

Of the 27 relapse events in total, clustering features were missing before 3 (11%) events owing to missing signal modalities. For 11 (46%) of the remaining 24 relapses, anomalies in clustering features were observed qualitatively in the time series of these features before and after the relapse. Most of these anomalies represented a transition to a cluster with inactive sensor readings; for example, GMM cluster 3 and PAM cluster 1 (Figure 5). We hypothesized that the patients, for whom we see their assigned cluster labels near relapse period being assigned to the cluster of inactive sensor recordings, most likely had their phone turned off a few days before the relapse. This transition to an inactive cluster was associated with an increase in likelihood scores (GMM-based feature) and a decrease in distance scores (PAM model-based feature) because these clusters were more compact and points did not deviate very much from the cluster centers.

Our cluster analysis between the NRx and pre-NRx periods showed that, on average, likelihood scores increased and distance scores decreased closer to relapses (Figure 6). This trend was robust with respect to different window sizes, and the largest change was observed with NR20 window size. Asterisks indicate that the absolute Cliff δ value between the 2 periods is >0.147 (ie, effect is nonnegligible; refer to *Analyzing Cluster Results* in the *Methods* section). Note that the plots were made using all patients' data collectively. Individual patient's plots would show a large difference between the near-relapse window and healthy period. Average Cliff δ values across all relapse events are presented in Table S1 in Multimedia Appendix 1.

We evaluated the relapse prediction pipeline discussed in *Relapse Prediction* in the *Methods* section, with and without the clustering-based features. The highest F2 score of 0.23 was obtained when the baseline features were complemented with the clustering-based features, which was significantly higher than the F2 score from random classification baseline (0.042) and the F2 score obtained using the baseline features only (0.18).

Figure 5. Time series plots of cluster assignment as obtained from the Gaussian mixture model (GMM) and partition around medoids (PAM) models (left pane) and weighted average likelihood score and distance score of a sample patient (right pane). Changes in cluster features are seen near the relapse instance (shown with the vertical red line).

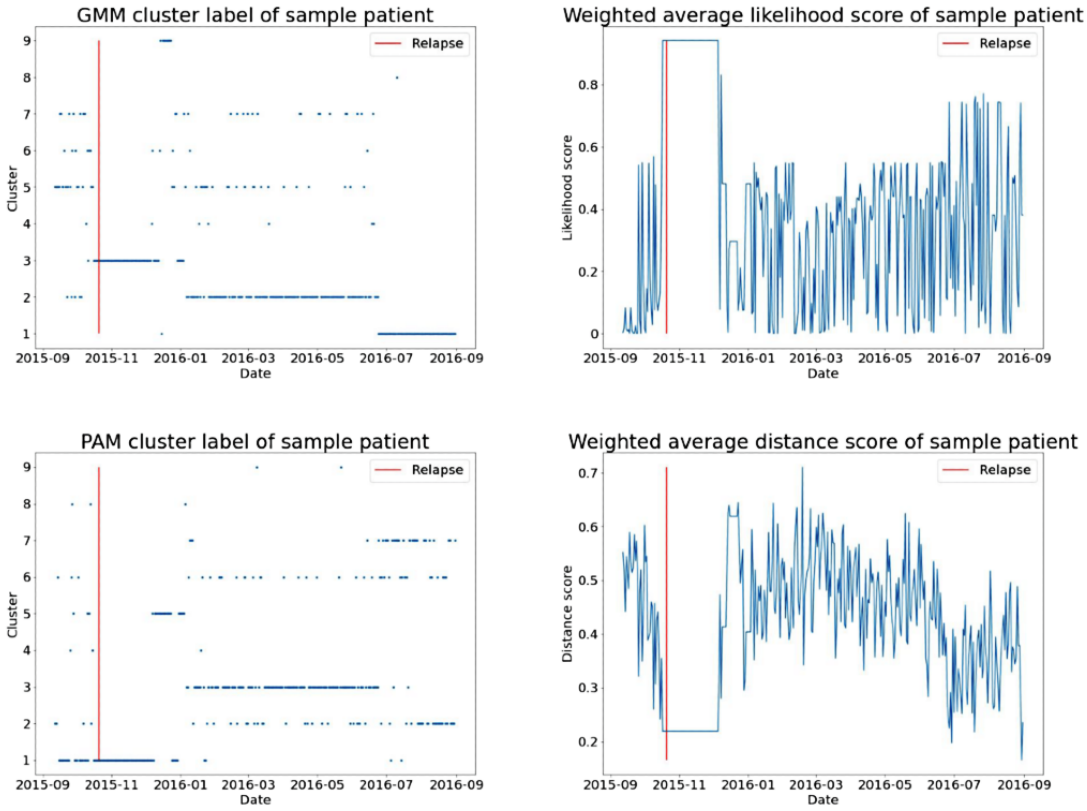
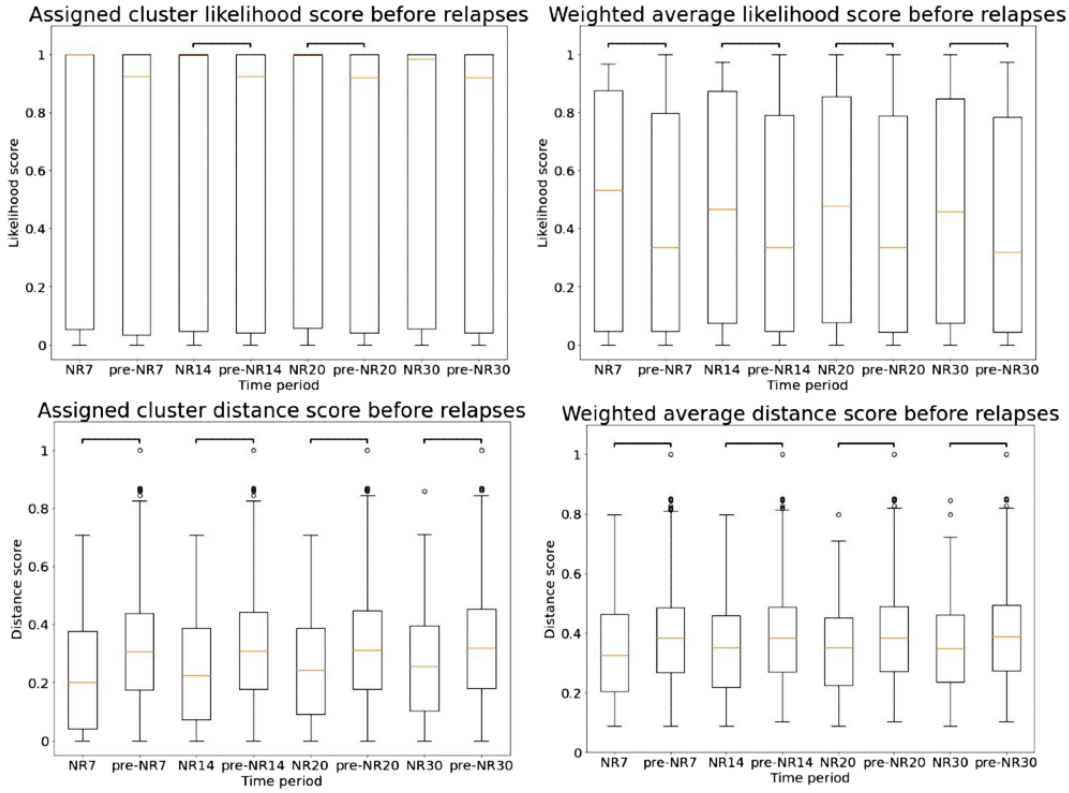


Figure 6. Boxplot of the clustering features (likelihood scores from Gaussian mixture model on top and distance scores from partition around medoids model at the bottom) in x days near relapse (NRx) and all days before relapses not in NRx (pre-NRx) periods. Bars indicate nonnegligible effect size.



Significant Features

With the best relapse prediction obtained using all features, we identified the most important features within this feature set based on how often a feature was selected in the leave-one-patient-out cross-validation. The selection count for a feature was incremented by 1 if it was selected for use in a particular cross-validation loop for a test patient. It should be noted that the number of features selected in each cross-validation loop was different because the number of features was a hyperparameter selected using nested cross-validation. Then, we normalized the total selection count of each feature at the end of the cross-validation by the number of cross-validation loops. The results obtained are presented in Table 3.

Table 3. The top 10 significant features in the relapse prediction pipeline based on the entire feature set (baseline and clustering-based features). The frequency of selection of a particular feature across the cross-validation loop is used to assess the most significant features for relapse prediction. It is to be noted that different numbers of features are selected in each cross-validation loop, as the number of features to be used is a hyperparameter tuned with a nested cross-validation loop.

Features	Frequency (normalized)
Baseline feature–distance template skewness	0.19
Clustering feature–mean PAM[a] label	0.17
Clustering feature–mean PAM weighted distance	0.14
Baseline feature–conversation template skewness	0.14
Clustering feature–number of transitions	0.12
Clustering feature–SD GMM[b] label	0.10
Clustering feature–SD PAM label	0.10
Clustering feature–mean GMM assigned cluster likelihood	0.10
Baseline feature–conversation template kurtosis	0.08
Baseline feature–volume template range	0.08

[a]PAM: partition around medoids.

[b]GMM: Gaussian mixture model.

Discussion

Principal Findings

In this study, we used clustering models to obtain behavioral representation from mobile sensing data, which could be useful for relapse prediction. The two clustering models explored in this study, GMM and PAM, grouped observations using different notions of distance or similarity between data points, and therefore, captured different behavioral representations (Table 2; Figure 4). These representations can be useful in downstream applications such as relapse prediction.

The GMM defines distance based on one-to-one matching between the hourly observation of mobile sensing data in the daily template. The clusters identified from the GMM have a widely varied distribution of cluster spread (Figure 3). With some compact clusters (represented by low cluster covariance) being identified within the GMM, the remaining data points that do not belong to any of these compact clusters are considered as a large-spread cluster with no typical cluster profile. These large-spread clusters also contain the compact clusters (cluster overlaps); a point belonging to compact clusters also shows high likelihood of belonging to the large-spread cluster. As we wanted the clusters to capture distinct behavioral trends, we evaluated Bhattacharyya distances to identify the best clustering model with least overlap between identified clusters. The PAM model with DTW distance allows a more lenient match of daily templates of behaviors as represented by the mobile sensing-based features. Such a lenient matching fits the context of this study because DTW is able to adjust for spikes, speed differences, or time shifts when evaluating dissimilarity between 2 daily templates of behaviors. However, the clusters obtained from the PAM model contain more dissimilarities. Then, it is more difficult to summarize the cluster profiles for qualitative model interpretation.

Overall, GMM-based modeling can identify highly dense or populous clusters with very specific behaviors associated with these clusters and some dispersed clusters that do not have a typical cluster profile. For example, cluster 3 and cluster 9 identified from the GMM (Table 2) represented 2 types of typical routines. Cluster 3 from the GMM had almost all sensor readings, except sleep, close to 0, likely representing an inactive or sedentary day, and cluster 9 had the days with the phone screen always turned on, likely representing a day with high mobile phone use. The PAM model also had a cluster with mostly inactive days and constantly long screen time (cluster 5). However, this cluster had higher cluster variance. When the average cluster profile of this cluster was observed (Figure 4), some days that did not strictly follow the patterns of inactive day and long screen time were also assigned to the cluster. In terms of behavioral features, this implies that clusters obtained from a PAM model are likely to cluster together the behaviors that do not always show homogeneity based on qualitative observations. This is because of the flexibility of the PAM model in allowing unparalleled alignment between behavior time series. Nonetheless, it might be beneficial to consider PAM-based modeling for the previously mentioned features: ability to

discount spikes and speed differences or time shifts when evaluating dissimilarity between 2 daily templates of behaviors. Similarity (or dissimilarity) between behaviors may not always be fully represented by hourly alignment and comparison of mobile sensing data across days.

The behavior of a particular day, represented by the mobile sensing data template for that day, was characterized in a clustering model with different clustering-based features such as Gaussian likelihood and DTW distance to the cluster centers. Days with assigned cluster likelihood scores close to 1 and assigned cluster distance scores close to 0 tend to belong to a dense cluster with a small spread. For example, cluster 3 from the GMM and cluster 5 from the PAM model have the highest likelihood and lowest distance to its assigned cluster, respectively (Figure S4 in Multimedia Appendix 1). They also have low within-cluster variability, as measured by the trace of sample covariance (Figure 3). On the other hand, days characterized by low likelihood scores and high distance scores tend to be more dispersed and do not conform well to a specific routine. For example, cluster 5 and cluster 7 in GMM and cluster 9 in the PAM model have such properties. Overall, GMM clustering and PAM clustering tend to produce clusters with different behavioral representations in the assigned clusters, and this is reflected in the clustering-based features such as likelihood scores and cluster distance that are assigned to characterize each day.

In terms of relapse prediction, clustering-based features can capture long-term behavioral trends across the patients. This representation can complement existing approaches to behavioral representation for psychotic relapse prediction in schizophrenia; for example, based on the use of daily behavioral rhythm change features as proposed by Lamichhane et al [19]. We compared the clustering-based features before and near the relapse periods and observed significant differences in some features. This was also seen qualitatively in a time series plot of clustering-based features, indicating that an upcoming relapse for a patient is associated with changes seen in clustering-based features (Figure 5). Clustering-based features were helpful in relapse prediction models (Table 4).

When clustering-based features were used together with daily behavioral rhythm change features, a significant gain in relapse prediction performance was obtained (F2 score improved from 0.18 to 0.23). These F2 scores and the associated improvements are significant, considering that a random classification baseline gives an F2 score of 0.042 on average. A Wilcoxon signed-rank test on performances in multiple classifier initializations for classification with and without clustering features yielded a significantly high score for classification when clustering features were included (P=.002). Clustering-based features were among the top features when significance of features for the relapse prediction task was evaluated (Table 3). Features such as mean cluster labels and number of transitions of labels were among the top (most frequently selected) features. Thus, both the information about which behavioral clusters the observations from the current period of monitoring belong to (likely representing behavioral clusters that are not normal behaviors) and how often transitions between different behavioral clusters occur (representing the patient showing frequent behavioral variations) are likely predictive of an oncoming relapse. Clustering-based features alone also proved to be valuable for relapse prediction. GMM-based and PAM-based clustering features only used in the relapse prediction pipeline led to an F2 score of 0.16 and 0.16 for relapse prediction, respectively (Table 4). Therefore, clustering-based features are found to be a useful approach to obtain behavioral representations and can be used in clinical applications such as relapse prediction.

Table 4. Relapse prediction performance with different feature sets. The baseline features introduced in the previous study [19] are complemented with clustering-based features for evaluation. The performance of both the GMM[a]-based and PAM[b]-based feature sets are also separately evaluatedc.

Method	F2 score (precision/recall)
All features	0.23 (0.063/0.662)
Baseline features [19]	0.18 (0.055/0.400)
Clustering features	0.14 (0.035/0.487)
GMM features	0.16 (0.042/0.487)
PAM features	0.19 (0.042/0.525)
GMM+baseline features	0.19 (0.052/0.525)
PAM+baseline features	0.16 (0.045/0.438)

[a]GMM: Gaussian mixture model.

[b]PAM: partition around medoids.

[c]Random classification baseline: mean score 0.042 (SD 0.020).

Comparison With Previous Studies, Limitations, and Future Research

To the best of our knowledge, this is the first study that used clustering analysis to group behavioral patterns of individuals with schizophrenia. Compared with previous studies that used hourly data to train the relapse prediction models, our study, based on clustering features to represent different behavioral patterns, has better model interpretability. Clustering analysis allows clinicians to understand different types of patient routines and their frequencies. In terms of schizophrenia, cluster transitions observed before relapses could suggest which types of behavior are potential relapse-related behavioral signatures.

Then, intervention strategies to prevent relapses can be made accordingly.

Researchers have studied how missing data are related to relapses and anomalies in mental health conditions. In the data set that we used for evaluation in this study, some passive-sensing daily templates have consecutive hours with missing data from almost all signal modalities. In an anomaly detection study, Adler et al [21] used mean imputation, whereas here, we filled missing values with zeros. Filling missing data with mean values may ignore the potential relationship between missing data and anomalies. In reality, it is highly possible that outpatients may turn off their phone when they experience relapse symptoms. We observed that there are more days from an inactive sensor reading cluster closer to relapses. The increased prevalence of inactive days also caused likelihood scores to increase and distance scores to decrease before relapses. Initially, we hypothesized that adhering to any routine or cluster center might reduce the risk of relapse, but it turned out that some routines, such as missing sensing data, are actually associated with a higher risk of relapse.

Although the clustering features successfully improved relapse prediction results, the only observable relapse signature was an increase in likelihood score or a decrease in distance score and a transition to an inactive cluster. For the relapse events that were not indicated by sensor inactiveness, we did not find any nontrivial changes in any specific feature before the relapse. Similarly, the relapse prediction performance with the best F2 score of 0.23 is relatively low. However, investigations of mobile sensing-based relapse prediction in mental health disorders are relatively new, and further improvements in this field can be expected as more data sets become available and improvements in machine learning models for the specific task of relapse prediction are made. In the study by Borges et al [45], relapse prediction in bipolar disorder was developed using clinical assessment features during patient visits. A high F score (F1) of up to 0.80 was reported. The relapse rate was quite high (relapse in >60% of the included patients) in the data set used by the authors, and the relapse prediction was performed at a patient level (instead of a weekly prediction model in free-living conditions, as in our case), which might have led to high performance.

In this study, we obtained patient-independent clusters; that is, generalized behavioral clusters, by pooling data from all the patients. We generalized that there are certain types of routines across all outpatients with schizophrenia. Future studies can focus on establishing personalized cluster models. As suggested in the study by He-Yueya et al [25], every patient's relapse signatures and the extent to which they adhere to their daily routine are different. The same study found that individual-level models can achieve better performance in predicting symptom severity. Our model also found that participants have different routines, as their frequencies of staying in different clusters largely varies. Moreover, although most patients had high likelihood scores and low distance scores closer to relapses, some other patients demonstrated the opposite trend. Generalized behavioral models might not fully represent and discount the effect of different confounding variables such as job type, sex, current health, and so on, which could impact behavioral trends. Although we used model personalization in relapse prediction, we considered only age as a covariate of behavioral trends. Personalized cluster models that account for different aspects of interpersonal differences would further help mitigate possible biases in behavioral representations owing to confounding variables. Personalized relapse prediction models will also be required to test the effectiveness of the individual-level clusters. However, sufficient data for each new patient are needed to find cluster models specific to the patient, and thus, clinical deployment for new patients will be delayed. Cluster adaptation from generalized cluster models to personalized cluster models, as more patient-specific data become available, needs to be investigated in future studies.

Conclusions

In this study, we proposed a methodology to compute clustering models on 24-hour daily behavior of outpatients with schizophrenia and showed that information extracted from the cluster model improved relapse prediction. New features were generated from the cluster models by measuring the deviation of every observation from the cluster centers representing typical behavioral patterns. Two different clustering models were investigated. The GMM allows for cluster overlap and has a more extreme cluster dispersion. The PAM model with DTW distance creates partitional clusters that are more generalized toward new data, but fails to identify dense clusters. The clustering-based features helped to improve relapse prediction model performance. In future studies, we will further investigate personalized clusters and relapse prediction models.

Acknowledgments

The authors appreciate all the patients who joined the study, the study team, and the support from the National Institute of Health's Exceptional Unconventional Research Enabling Knowledge Acceleration program (R01MH103148) and National Science Foundation grants (#1840167 and #2047296).

Conflicts of Interest

DBZ has financial interests in Merlin LLC, Focus technology, and Core technology. He has an intervention content licensing agreement with Pear Therapeutics and has provided consultation services to Trusst Health, K Health, Boehringer Ingelheim, eQuility, Deep Valley Labs, and Otsuka Pharmaceuticals. AS has received travel reimbursement or honorarium payments from Gordon Research Conferences, Pola Chemical Industries, Leuven Mindgate, American Epilepsy Society, and IEEE. AS has

received research support from Microsoft, Sony Corporation, NEC Corporation, and Pola Chemicals and consulting fees from Gideon Health and Suntory Global Innovation Center. AS was paid by the European Science Foundation for a grant review.

Multimedia Appendix 1
Additional information, figures, and tables of clustering model selection, relapse prediction model selection, clustering, and relapse prediction results.
[PDF File (Adobe PDF File), 295 KB - mhealth_v10i4e31006_app1.pdf]

References

1. GBD 2017 DiseaseInjury IncidencePrevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1789-1858 [FREE Full text] [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
2. Charlson F, Ferrari A, Santomauro D, Diminic S, Stockings E, Scott JG, et al. Global epidemiology and burden of schizophrenia: findings from the global burden of disease study 2016. Schizophr Bull 2018 Oct 17;44(6):1195-1203 [FREE Full text] [doi: 10.1093/schbul/sby058] [Medline: 29762765]
3. Smith T, Weston C, Lieberman J. Schizophrenia (maintenance treatment). Am Fam Physician 2010 Aug 15;82(4):338-339 [FREE Full text] [Medline: 20704164]
4. Diagnostic And Statistical Manual Of Mental Disorders, Fifth Edition. Washington, D.C., United States: American Psychiatric Association; 2013.
5. Olfson M, Gerhard T, Huang C, Crystal S, Stroup T. Premature mortality among adults with schizophrenia in the United States. JAMA Psychiatry 2015 Dec;72(12):1172-1181 [FREE Full text] [doi: 10.1001/jamapsychiatry.2015.1737] [Medline: 26509694]
6. Emsley R, Chiliza B, Asmal L, Harvey B. The nature of relapse in schizophrenia. BMC Psychiatry 2013 Feb 8;13(1):50 [FREE Full text] [doi: 10.1186/1471-244x-13-50]
7. Robinson D, Woerner M, Alvir J, Bilder R, Goldman R, Geisler S, et al. Predictors of relapse following response from a first episode of schizophrenia or schizoaffective disorder. Arch Gen Psychiatry 1999 Mar;56(3):241-247 [FREE Full text] [doi: 10.1001/archpsyc.56.3.241] [Medline: 10078501]
8. Kane J. Treatment strategies to prevent relapse and encourage remission. J Clin Psychiatry 2007;68 Suppl 14:27-30. [Medline: 18284275]
9. Overall J, Gorham D. The brief psychiatric rating scale. Psychol Rep 2016 Aug 31;10(3):799-812 [FREE Full text] [doi: 10.2466/pr0.1962.10.3.799]
10. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, et al. Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2014 Presented at: UbiComp '14: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing; Sep 13 - 17, 2014; Seattle Washington URL: https://doi.org/10.1145/2632048.2632054 [doi: 10.1145/2632048.2632054]
11. Saeb S, Zhang M, Karr C, Schueller S, Corden M, Kording K, et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. J Med Internet Res 2015 Jul 15;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
12. Ben-Zeev D, Wang R, Abdullah S, Brian R, Scherer E, Mistler L, et al. Mobile behavioral sensing for outpatients and inpatients with schizophrenia. Psychiatr Serv 2016 May 01;67(5):558-561 [FREE Full text] [doi: 10.1176/appi.ps.201500130] [Medline: 26695497]
13. Shin S, Yeom C, Shin C, Shin J, Jeong J, Shin J, et al. Activity monitoring using a mHealth device and correlations with psychopathology in patients with chronic schizophrenia. Psychiatry Res 2016 Dec 30;246:712-718 [FREE Full text] [doi: 10.1016/j.psychres.2016.10.059] [Medline: 27836243]
14. Cella M, Okruszek L, Lawrence M, Zarlenga V, He Z, Wykes T. Using wearable technology to detect the autonomic signature of illness severity in schizophrenia. Schizophr Res 2018 May;195:537-542 [FREE Full text] [doi: 10.1016/j.schres.2017.09.028] [Medline: 28986005]
15. Tron T, Resheff Y, Bazhmin M, Peled A, Weinshall D. Real-time schizophrenia monitoring using wearable motion sensitive devices. In: Proceedings of the 26th European Congress of Psychiatry. 2018 Presented at: The 26th European Congress of Psychiatry; Mar 3-6, 2018; Nice URL: https://doi.org/10.1007/978-3-319-98551-0_28 [doi: 10.26226/morressier.5a7070e5d462b80290b580d1]
16. Gay K, Torous J, Joseph A, Pandya A, Duckworth K. Digital technology use among individuals with schizophrenia: results of an online survey. JMIR Ment Health 2016 May 04;3(2):e15 [FREE Full text] [doi: 10.2196/mental.5379] [Medline: 27146094]

17. Barnett I, Torous J, Staples P, Sandoval L, Keshavan M, Onnela J. Relapse prediction in schizophrenia through digital phenotyping: a pilot study. Neuropsychopharmacology 2018 Jul;43(8):1660-1666 [FREE Full text] [doi: 10.1038/s41386-018-0030-z] [Medline: 29511333]
18. Wang R, Wang W, Obuchi M, Scherer E, Brian R, Ben-Zeev D, et al. On predicting relapse in schizophrenia using mobile sensing in a randomized control trial. In: Proceedings of the 2020 IEEE International Conference on Pervasive Computing and Communications (PerCom). 2020 Presented at: 2020 IEEE International Conference on Pervasive Computing and Communications (PerCom); Mar 23-27, 2020; Austin, TX, USA URL: https://doi.org/10.1109/percom45495.2020.9127365 [doi: 10.1109/percom45495.2020.9127365]
19. Lamichhane B, Ben-Zeev D, Campbell A, Choudhury T, Hauser M, Kane J, et al. Patient-independent schizophrenia relapse prediction using mobile sensor based daily behavioral rhythm changes. In: Wireless Mobile Communication and Healthcare. Cham: Springer; 2021.
20. Ben-Zeev D, Brian R, Wang R, Wang W, Campbell A, Aung M, et al. CrossCheck: integrating self-report, behavioral sensing, and smartphone use to identify digital indicators of psychotic relapse. Psychiatr Rehabil J 2017 Sep;40(3):266-275 [FREE Full text] [doi: 10.1037/prj0000243] [Medline: 28368138]
21. Adler D, Ben-Zeev D, Tseng V, Kane J, Brian R, Campbell A, et al. Predicting early warning signs of psychotic relapse from passive sensing data: an approach using encoder-decoder neural networks. JMIR Mhealth Uhealth 2020 Aug 31;8(8):e19962 [FREE Full text] [doi: 10.2196/19962] [Medline: 32865506]
22. Lin E, Weintraub M, Miklowitz D, Chen P, Lee S, Chen H, et al. The associations between illness perceptions and social rhythm stability on mood symptoms among patients with bipolar disorder. J Affect Disord 2020 Aug 01;273:517-523 [FREE Full text] [doi: 10.1016/j.jad.2020.05.019] [Medline: 32560948]
23. Bond M, Feather N. Some correlates of structure and purpose in the use of time. J Personality Soc Psychol 1988;55(2):321-329 [FREE Full text] [doi: 10.1037/0022-3514.55.2.321]
24. Tseng V, Sano A, Ben-Zeev D, Brian R, Campbell A, Hauser M, et al. Using behavioral rhythms and multi-task learning to predict fine-grained symptoms of schizophrenia. Sci Rep 2020 Sep 15;10(1):15100 [FREE Full text] [doi: 10.1038/s41598-020-71689-1] [Medline: 32934246]
25. He-Yueya J, Buck B, Campbell A, Choudhury T, Kane J, Ben-Zeev D, et al. Assessing the relationship between routine and schizophrenia symptoms with passively sensed measures of behavioral stability. NPJ Schizophr 2020 Nov 23;6(1):35 [FREE Full text] [doi: 10.1038/s41537-020-00123-2] [Medline: 33230099]
26. Wang R, Aung M, Abdullah S, Brian R, Campbell A, Choudhury T, et al. CrossCheck: toward passive sensing and detection of mental health changes in people with schizophrenia. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2016 Presented at: UbiComp '16: The 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing; Sep 12 - 16, 2016; Heidelberg Germany URL: https://doi.org/10.1145/2971648.2971740 [doi: 10.1145/2971648.2971740]
27. A new paradigm for illness monitoring and relapse prevention in schizophrenia. ClinicalTrials. URL: https://clinicaltrials.gov/show/NCT019520412013 [accessed 2022-03-10]
28. Wang R, Wang W, Aung MS, Ben-Zeev D, Brian R, Campbell AT, et al. Predicting symptom trajectories of schizophrenia using mobile sensing. Proc ACM Interact Mob Wearable Ubiquitous Technol 2017 Sep 11;1(3):1-24 [FREE Full text] [doi: 10.1145/3130976]
29. Buck B, Scherer E, Brian R, Wang R, Wang W, Campbell A, et al. Relationships between smartphone social behavior and relapse in schizophrenia: a preliminary report. Schizophr Res 2019 Jun;208:167-172 [FREE Full text] [doi: 10.1016/j.schres.2019.03.014] [Medline: 30940400]
30. Geary D, McLachlan G, Basford K. Mixture models: inference and applications to clustering. J Royal Stat Soc Series A (Statistics in Society) 1989;152(1):126 [FREE Full text] [doi: 10.2307/2982840]
31. Haben S, Singleton C, Grindrod P. Analysis and clustering of residential customers energy behavioral demand using smart meter data. IEEE Trans Smart Grid 2016 Jan;7(1):136-144 [FREE Full text] [doi: 10.1109/tsg.2015.2409786]
32. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. JMLR 2011;12(85):2825-2830 [FREE Full text]
33. Dempster A, Laird N, Rubin D. Maximum likelihood from incomplete data via the algorithm. J Royal Stat Soc Series B (Methodological) 2018 Dec 05;39(1):1-22 [FREE Full text] [doi: 10.1111/j.2517-6161.1977.tb01600.x]
34. Kailath T. The divergence and bhattacharyya distance measures in signal selection. IEEE Trans Commun 1967 Feb;15(1):52-60 [FREE Full text] [doi: 10.1109/tcom.1967.1089532]
35. Arthur D, Vassilvitskii S. K-means++: the advantages of careful seeding. In: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms. 2007 Presented at: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms; Jan 7 - 9, 2007; New Orleans Louisiana. [doi: 10.5555/1283383.1283494]
36. De Maesschalck R, Jouan-Rimbaud D, Massart D. The Mahalanobis distance. Chemometrics Intell Laboratory Syst 2000 Jan;50(1):1-18 [FREE Full text] [doi: 10.1016/s0169-7439(99)00047-7]
37. Waterman M. Time warps, string edits, and macromolecules: the theory and practice of sequence comparison. Mathematical Biosciences 1985 Oct;76(2):243-244 [FREE Full text] [doi: 10.1016/0025-5564(85)90007-0]

38. Kaufman L, Rousseeuw PJ. Clustering by means of mediods. Kuleuven. URL: https://wis.kuleuven.be/stat/robust/papers/publications-1987/kaufmanrousseeuw-clusteringbymedoids-l1norm-1987.pdf [accessed 2022-03-10]
39. Cliff N. Dominance statistics: ordinal analyses to answer ordinal questions. Psychol Bulletin 1993;114(3):494-509 [FREE Full text] [doi: 10.1037/0033-2909.114.3.494]
40. Romano J, Kromrey J, Coraggio J, Skowronek J. Appropriate statistics for ordinal level data: should we really be using t-test and Cohen's d for evaluating group differences on the NSSE and other surveys? BibSonomy. URL: https://www.bibsonomy.org/bibtex/216a5c27e770147e5796719fc6b68547d/kweiand [accessed 2022-03-10]
41. Rui W. Mental health sensing using mobile phones. Dartmouth College. URL: https://www.cs.dartmouth.edu/~campbell/dr_rui_wang_thesis-2018.pdf; [accessed 2022-03-10]
42. Lemaître G, Nogueira F, Aridas C. Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. J Mach Learn Res 2017;18(1):559-563. [doi: 10.5555/3122009.3122026]
43. Alphs L, Nasrallah H, Bossie C, Fu D, Gopal S, Hough D, et al. Factors associated with relapse in schizophrenia despite adherence to long-acting injectable antipsychotic therapy. Int Clin Psychopharmacol 2016;31(4):202-209 [FREE Full text] [doi: 10.1097/yic.0000000000000125]
44. Jeste DV, Maglione JE. Treating older adults with schizophrenia: challenges and opportunities. Schizophr Bull 2013 Sep;39(5):966-968 [FREE Full text] [doi: 10.1093/schbul/sbt043] [Medline: 23552180]
45. Borges R, Salvini R, Nierenberg A, Sachs G, Lafer B, Dias R. Forecasting depressive relapse in Bipolar Disorder from clinical data. In: Proceedings of the 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2018 Presented at: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); Dec 3-6, 2018; Madrid, Spain URL: https://doi.org/10.1109/bibm.2018.8621255 [doi: 10.1109/bibm.2018.8621255]

Abbreviations

AIC: Akaike information criterion
BIC: Bayesian information criterion
BRF: balanced random forest
DTW: dynamic time warping
GMM: Gaussian mixture model
NRx: x days near relapse period
PAM: partition around medoids

Edited by L Buis; submitted 12.06.21; peer-reviewed by J Ainsworth, D Sato, B Booth, A Hudon; comments to author 29.07.21; revised version received 19.10.21; accepted 17.02.22; published 11.04.22.

Please cite as:
Zhou J, Lamichhane B, Ben-Zeev D, Campbell A, Sano A
Predicting Psychotic Relapse in Schizophrenia With Mobile Sensor Data: Routine Cluster Analysis
JMIR Mhealth Uhealth 2022;10(4):e31006
URL: https://mhealth.jmir.org/2022/4/e31006
doi:10.2196/31006
PMID:35404256

Original Paper

Low- and High-Intensity Physical Activity Among People with HIV: Multilevel Modeling Analysis Using Sensor- and Survey-Based Predictors

Paul Cook[1], PhD; Catherine Jankowski[1], PhD; Kristine M Erlandson[2], MD; Blaine Reeder[3], PhD; Whitney Starr[2], MSN; Mary Beth Flynn Makic[1], PhD

[1]College of Nursing, University of Colorado, Aurora, CO, United States

[2]School of Medicine, University of Colorado, Aurora, CO, United States

[3]Sinclair School of Nursing, University of Missouri, Columbia, MO, United States

Corresponding Author:
Paul Cook, PhD
College of Nursing
University of Colorado
13120 E 19th Ave
Campus Box C288-04
Aurora, CO, 80045
United States
Phone: 1 3037248537
Fax: 1 3037248560
Email: paul.cook@cuanschutz.edu

Abstract

Background: High-intensity physical activity improves the health of people with HIV. Even when people have good intentions to engage in physical activity, they often find it difficult to maintain physical activity behavior in the long term. Two Minds Theory is a neurocognitive model that explains gaps between people's intentions and behaviors based on the operations of 2 independent mental systems. This model predicts that everyday experiences will affect physical activity and that factors outside people's awareness, such as sleep and stress, can have particularly strong effects on physical activity behaviors.

Objective: We designed this study to test the effects of daily experiences on physical activity among people with HIV, including measures of people's conscious experiences using daily electronic surveys and measures of nonconscious influences using sensor devices.

Methods: In this study, 55 people with HIV wore a Fitbit Alta for 30 days to monitor their physical activity, sleep, and heart rate variability (HRV) as a physiological indicator of stress. Participants also used their smartphones to complete daily electronic surveys for the same 30 days about fatigue, self-efficacy, mood, stress, coping, motivation, and barriers to self-care. Time-lagged, within-person, multilevel models were used to identify the best prospective predictors of physical activity, considering the daily survey responses of people with HIV and sensor data as predictors of their physical activity the following day. We also tested baseline surveys as predictors of physical activity for comparison with daily variables.

Results: Different people had different average levels of physical activity; however, physical activity also varied substantially from day to day, and daily measures were more predictive than baseline surveys. This suggests a chance to intervene based on day-to-day variations in physical activity. High-intensity physical activity was more likely when people with HIV reported less subjective fatigue on the prior day (r=−0.48) but was unrelated to actual sleep based on objective sensor data. High-intensity physical activity was also predicted by higher HRV (r=0.56), indicating less stress, lower HIV-related stigma (r=−0.21), fewer barriers to self-care (r=−0.34), and less approach coping (r=−0.34). Similar variables predicted lower-level physical activity measured based on the number of steps per day of people with HIV.

Conclusions: Some predictors of physical activity, such as HRV, were only apparent based on sensor data, whereas others, such as fatigue, could be measured via self-report. Findings about coping were unexpected; however, other findings were in line with the literature. This study extends our prior knowledge on physical activity by demonstrating a prospective effect of everyday experiences on physical activity behavior, which is in line with the predictions of Two Minds Theory. Clinicians can support the physical activity of people with HIV by helping their patients reduce their daily stress, fatigue, and barriers to self-care.

(JMIR Mhealth Uhealth 2022;10(4):e33938) doi:10.2196/33938

KEYWORDS

ecological momentary assessment; fatigue; HIV; physical activity; stress; mobile phone

Introduction

Background

Physical activity is important in managing many chronic diseases, including HIV. With current antiretroviral treatment (ART) options, people with HIV have almost but not quite a normal life expectancy [1]. Lingering disease-related morbidity and mortality among people with HIV are likely because of accelerated or accentuated aging [2], which may be tied to chronic inflammation and immune activation caused by the virus and by the immune system's efforts to attack it [3]. In particular, long-term HIV infection predisposes people to develop age-related comorbidities such as heart disease and diabetes at a younger age than their peers without HIV [2].

Benefits of Physical Activity for People With HIV

Assuming that a patient's HIV is controlled with ART, prevention of cardiovascular complications in people with HIV involves a number of health promotion strategies similar to those recommended for people without HIV. Hypertension, diabetes, and cholesterol screening are needed, and medications may be required to manage these conditions [4]. Smoking cessation should be discussed, and patients should be supported to quit smoking if desired [5]. Regular exercise, including cardiovascular, strength, flexibility, and balance training, is also recommended. People with HIV, as a group, have less physical activity than that recommended by public health guidelines [6]. In fact, an emerging line of research suggests that people with HIV need more intensive levels of exercise than usual to offset the negative cardiovascular effects of long-term HIV infection [7]. To achieve needed benefits for measures of physical function and cardiovascular fitness, people with HIV require high-intensity physical activity that elevates their heart rate (HR) for at least 30 minutes most days [8]. This is more intensive physical activity than is typical for most Americans [9].

In addition to its cardiovascular benefits, exercise is known to enhance mood and reduce stress, with higher levels of physical activity showing positive effects on many mental health indicators among people with HIV [10]. In addition, exercise interventions have been shown to improve neurocognitive function; improve mitochondrial function; reduce inflammation; improve bone mineral density; reduce fatigue; and, in some studies, directly improve immune function based on CD4 counts among people with HIV [11]. Physical activity may even improve ART medication adherence in people with HIV, an effect that seems to be mediated by reductions in depressive symptoms [12].

Barriers to Physical Activity Among People With HIV

Unfortunately, depressive symptoms make it less likely for people with HIV to engage in physical activity; other barriers to physical activity among people with HIV include ART side effects, comorbid health problems, physical pain, lower self-efficacy for exercise, fewer perceived benefits of exercise, and lower motivation for exercise [13]. Fatigue is the most common symptom experienced by people with HIV and is cited by patients as a barrier to exercise, although exercise actually tends to reduce fatigue [3]. Qualitative research suggests additional social or environmental barriers to physical activity, including concerns about HIV-related stigma, lack of interpersonal support for exercise, environmental or resource limitations that make physical activity more challenging, and difficulty incorporating exercise into daily activities [14]. Once an exercise habit is started, other barriers to maintenance may arise because the factors that help someone to initiate a physical activity habit are often different from those that help them to maintain it over time [15]; maintenance is more likely when people with HIV have social support and when exercise is well integrated into their daily routines [16].

Explaining Physical Activity via Two Minds Theory

The differences between the positive effects of exercise and people's difficulty in adopting and maintaining exercise routines can be explained by Two Minds Theory (TMT), which is a novel approach to understanding and changing health behaviors based on the idea of *intention-behavior gaps* [17]. TMT explains this discrepancy by positing that intentions and behaviors are produced by 2 separate neurocognitive systems. Behavior arises from the *intuitive system*, a set of deep brain structures and functions that perceives situations, compares them to memories, generates emotions, triggers behavioral responses, and is shaped by consequences. All of these intuitive-level processes can occur largely outside of people's conscious awareness. In contrast, intentions for the future (as well as explanations about past behaviors) are produced by the *narrative system*, a set of conscious higher brain processes coordinated by the prefrontal cortex that interprets and draws conclusions about a person's experiences and behaviors. The intuitive system is much more efficient and automatic but is more strongly affected by emotions, surface characteristics of a situation, and immediate barriers or rewards; thus, it is sometimes fooled or misled [18]. The narrative–intuitive distinction also helps to explain why, even after a person begins an exercise habit, unexpected barriers can interfere with maintaining it. In general, people start a new behavior because it seems important (ie, when narrative-level benefits are high); however, they continue the new behavior over time when it is easy (ie, when intuitive-level barriers such as HIV-related stigma and environmental challenges are low) [15].

As the intuitive system operates in the moment and in the context of everyday environments, Cook et al [17] suggested that researchers must examine people's immediate experiences to understand their behavior. Many of the barriers to physical activity for people with HIV fall under the heading of immediate experiences; negative mood, pain, fatigue, ART side effects, and low motivation are all experiential states that go up or down

over time for a given person living with HIV [19]. The best practice for studying in-the-moment states is ecological momentary assessment (EMA) using diaries or electronic surveys to collect real-time data on multiple occasions from the same person [20]. Some everyday states are not necessarily accessible for conscious reporting; for example, fatigue may be a symptom experience that people with HIV can report or may manifest via poor sleep, psychomotor slowing, or cognitive confusion [3]. In such cases, ambulatory monitoring data from personal sensor devices may provide additional information [21]. In prior research involving people with HIV, everyday situational factors such as mood, stress, stigma, fatigue, and self-care barriers predicted both ART adherence [19] and fatigue [22]; based on TMT, we expect that these variables would also predict another daily variable, physical activity. Furthermore, we predict that nonconscious influences such as sleep and stress would have particularly strong effects on the subsequent physical activity of people with HIV.

Purpose of This Study

To better understand the factors that facilitate or interfere with physical activity among people with HIV, we conducted a secondary analysis of data from a mixed method study of daily fatigue experiences among people with HIV [22]. The purpose of the current analysis is to understand the factors that affect everyday physical activity among people with HIV, as measured by 1 month of monitoring using a Fitbit digital sensor device (Fitbit Inc). Predictors of physical activity were assessed using daily sensor data and survey data collected using EMA methods. As EMA generates many days' worth of data from each individual participant, it permits reliable conclusions to be drawn from studies with smaller sample sizes; the effective *N* for such studies is in between the number of participants and the number of observations after statistically correcting for the dependency of observations within participants.

Methods

Recruitment

Recruitment took place in an infectious disease specialty clinic at an academic medical center in Denver, Colorado, United States, from September 2017 to November 2018. The clinic provides care for approximately 1850 people with HIV annually, 97% of whom are on ART, 91% of whom are virally suppressed, and 30% to 45% of whom have significant fatigue [7,23]. The clinic is funded through the Ryan White program and uses a medical home model with on-site services, including mental health, pharmacy, and medical case management. HIV care providers at the clinic screened the patients for potential study eligibility based on the criteria described in the following sections. When a patient was potentially interested, a research team member met with them in the clinic to obtain informed consent. In most cases, the baseline study procedures were completed immediately. In addition, flyers for the study were placed in the clinic waiting room and examination rooms, and patients could self-refer and schedule a time for consent and baseline measures.

The inclusion criteria were people with HIV with (1) well-controlled HIV (viral load<20 copies/mL) with current ART, (2) English language fluency, (3) age of 18 to 80 years, and (4) at least mild fatigue based on the Patient-Reported Outcomes Measurement Information System (PROMIS) 4-item screening tool [24]. The exclusion criteria were (1) serious substance abuse; (2) cognitive impairment; or (3) mental or physical illness that, in the referring clinician's judgment, precluded participation.

Procedure

Participants wore a Fitbit Alta HR sensor device for 30 days. The overall physical activity was measured based on the total steps per day, whereas high-intensity physical activity was measured based on the number of *active minutes* when the participant had an elevated HR during exercise. Additional data collected from the Fitbit wristband included sleep metrics, HR indicators, and HR variability (HRV) as a physiological measure of stress.

Participants also completed the validated EMA survey measures on mood, fatigue, self-efficacy, and other psychological variables by responding to a daily message on their smartphones with a link to a REDCap (Research Electronic Data Capture; Vanderbilt University) web-based data form. The EMA survey included questions about situational variables that might interfere with physical activity, such as travel, substance use, or medication side effects.

After consent was obtained, people with HIV completed baseline questionnaires on fatigue and other symptoms. They provided a release of information to extract demographic and clinical data from their electronic health records and received instructions on how to use the Fitbit device and web-based EMA surveys for the next 30 days. A link to complete the survey was delivered once per day at a random time, either by email or SMS text message (participant's choice). Participants returned after 1 month to complete additional questionnaires and provide a blood sample, and the first 25 participants completed a qualitative interview about their fatigue experiences and physical activity; these data have been presented elsewhere [22].

Measures

Baseline Data

Demographic data from patients' charts and intake paperwork included age, gender, race, ethnicity, and current employment status. Participants also completed baseline self-report measures, including PROMIS scales for fatigue, depression, applied cognition, sleep, and pain [24]; the HIV Quality of Life scale [25]; and a short form of the HIV Stigma Scale [26].

Ambulatory Sensors

The Fitbit wristbands collected continuous ambulatory data on physical activity (total steps and active minutes per day), sleep (time in bed, total sleep time, wake after sleep onset, sleep efficiency, and sleep stages), HR (average, resting, minimum, and maximum), and HRV, with each of the metrics calculated as daily averages for 1 month. Consumer-grade Fitbit monitors provide a balance of sensitivity and data collection efficiency for both activity and sleep data, showing valid step counts compared with research-grade devices and acceptable results for acquiring HRV in field settings [27], as well as high

interdevice reliability for various sleep metrics [28]. Fitbit uses a proprietary algorithm to calculate active minutes based on the ratio of current energy expenditure to the amount of energy typically expended by that person at rest [29]; when that ratio is ≥3, Fitbit's algorithm considers the person to be engaged in high-intensity physical activity [30]. The Fitbit data were retrieved from the manufacturer's consumer-facing data portal and through an application programming interface [31]. This site allowed the researchers to download participants' HR data for each minute of the day and then calculate the daily HRV averages for each participant. Participants were allowed to keep the Fitbit device at the conclusion of the study. Participants also used a Pillsy electronic pill bottle to monitor ART adherence (data not presented here) over the same 30 days.

Daily Surveys

EMA survey items included the 4-item PROMIS fatigue short form (Cronbach α=.95) [24]; a 4-item self-efficacy scale from the Diary of Ambulatory Behavioral States (DABS) (Cronbach α=.84) [32]; the 3-item DABS mood scale (Cronbach α=.85) [32]; a 6-item stress scale adapted from the Daily Hassles Scale (Cronbach α=.75) [19]; the 9-item Assessment of Daily Coping (Cronbach α=.83), with 2 subscales measuring approach versus avoidance coping [33]; the 2-item DABS social support scale (Cronbach α=.81) [32]; 3 items from the HIV stigma scale (Cronbach α=.77) [26]; the 7-item Herzog motivation scale (Cronbach α=.79) [34]; an ART adherence item from the AIDS Clinical Trials Group measure [35]; and a 9-item barriers to self-care scale developed in our prior daily survey research (Cronbach α=.66), which measured daily situational variables such as travel, substance use, and medication side effects [19]. All reliability statistics reported are from data in this study.

Statistical Analysis

Analysis Plan

To quantify the effects of intuitive-level daily variables versus those of narrative-level measures collected once at baseline, we first examined within-person (day by day) and between-person (stable over time) variation on each of the physical activity measures using graphical displays and intraclass correlation coefficients (*ICCs*). The squared *ICC* is similar to the coefficient of determination (R^2) in regression analysis and can be interpreted as the percentage of variability in the outcome measure that is stable at the level of the individual participant—in other words, the between-person level of variability [36]. Any remaining variability that is *not* accounted for based on the squared *ICC* is the variability that occurs within individuals over time. Some of this remaining variability can potentially be predicted using day-by-day predictor variables that also vary within individuals over time.

We evaluated the effects of stable between-person characteristics using Pearson *r* correlations between participants' baseline characteristics and each of the 2 physical activity variables. To identify significant correlations, each Pearson *r* was converted to a *t* test using the *r*-to-*t* formula based on sample size, with *P* values obtained from the *t* distribution.

We then tested the effects of within-person sensor and survey variables via within-person multilevel models using SPSS (version 27, IBM Corp). Our models used restricted maximum likelihood estimation to generate β coefficients, which compensated for missing data by imputing a distribution within each participant's scores based on all available data points. The actual number of survey data points was 58.28% (or 943/1618 person-days) of the possible days. Of the 1618 person-days, data completeness for sensor measures varied by type, with 1130 (69.84%) person-days of valid HR and physical activity data but only 890 (55.01%) person-days of sensor-based sleep data.

A time-lagged analysis was used to predict each day's physical activity from the experiences of people with HIV on the previous day, which allows for causal conclusions based on the assumption that the cause precedes the effect [23]. Each predictor's effect on each of the two outcome variables—steps per day as a measure of low- to moderate-intensity exercise and number of active minutes as a measure of high-intensity exercise—was tested separately. For each outcome variable, each possible predictor was tested individually using a minimally restrictive .05 α level. All variables that passed this screening step were then combined into a single multivariable model to identify the most parsimonious set of predictors that together accounted for daily variations in physical activity among people with HIV. In the multivariable models, predictors were entered in block order (HR metrics, then sleep metrics, and then survey data), and backward elimination was used to remove redundant predictors within each block and in the final model.

Statistical Power

All models used fixed effects, a standard diagonal matrix, and group mean centering of predictors to control for the clustering of observations within participants. Under these assumptions, a sample size of 55 people with HIV yielded a power of 0.80 to detect moderate effect sizes (β>.40) in multilevel analyses, assuming up to 5 predictors with moderate multicollinearity (variance inflation factor 2.0), moderate *ICC* of 0.50, and α=.05.

Ethical Considerations

This study was approved by the Colorado Multiple Institutional Review Board (protocol 16-2603).

Results

Participant Characteristics

Of the 61 people with HIV approached for the study, 55 (90%) agreed to participate. The most common reasons for nonparticipation were (1) too busy for daily surveys, (2) not interested in using sensor devices, or (3) their smartphones being too old or low on memory to add the software needed for the study. The final sample included 85% (47/55) men and 15% (8/55) women. Participants' age ranged from 20 to 69 years, and 58% (32/55) of the participants were White and non-Hispanic. In addition, the sample included 20% (11/55) of Black participants and 22% (12/55) of Latino/Latina participants. These demographics are typical of people with HIV in Colorado.

Descriptive analyses of the baseline variables showed high levels of fatigue (greater than the PROMIS 50th percentile on 66% of days; 622/943), a high level of perceived stigma related to HIV (78% of days; 718/921), poor mood (at least once for 80% of people with HIV; 42/52), and high stress (at least once for 53% of people with HIV; 28/52). Furthermore, people with HIV frequently had interrupted sleep based on Fitbit data (284/888, 32% of nights), and some (8/40, 20%) had an overall sleep efficiency level <85%, suggesting disordered sleep. Consistent with the idea that even well-managed HIV involves chronic inflammation, participants had a moderate average C-reactive protein level (mean 2.51, SD 3.34 mg/L), and 5% (2/42) had high levels of inflammation based on C-reactive protein >10. Participant demographics were described in greater detail by Makic et al [22].

Patterns of Physical Activity

Physical activity varied dramatically both between and within individuals. Figure 1 shows the average activity levels (black dots) and SDs (gray bars) of the steps-per-day measurements for each participant. As shown in Figure 1, some people with HIV had higher average levels of physical activity than others; however, there was also a high level of within-person variability in physical activity from day to day. The heavy dashed line on the graph illustrates the recommended daily amount of overall physical activity (10,000 steps) for adults; in our sample, 61% (25/44) of participants either met or came close to meeting that goal, which suggests a higher level of physical activity among these people with HIV than among US adults overall. Some individual participants were very active, with an average of ≥15,000 steps per day. The relative length of the bars suggests that participants with lower activity levels were more likely to be consistently inactive; however, among those with high physical activity, there tended to be a broader range of activity from day to day. Overall, the *ICC* for this measure was 0.70, and 51% of the variance ($1 - ICC^2$) was not explained by a person's overall tendency toward activity or inactivity. In this context, we looked for within-person predictors that could help explain why some people with HIV exercised more on some days than others.

Figure 1. Steps per day for each study participant. Black dots represent each person's average steps across all days that they wore the Fitbit sensor device (up to 1 month for each participant). Gray bars represent the within-person average (SD 1). SDs were calculated within individuals, and therefore, the varying height of the bars reflects the fact that some people's daily amount of physical activity varied more than that of others. The dashed line at 10,000 steps reflects the recommended daily amount of overall physical activity for adults.

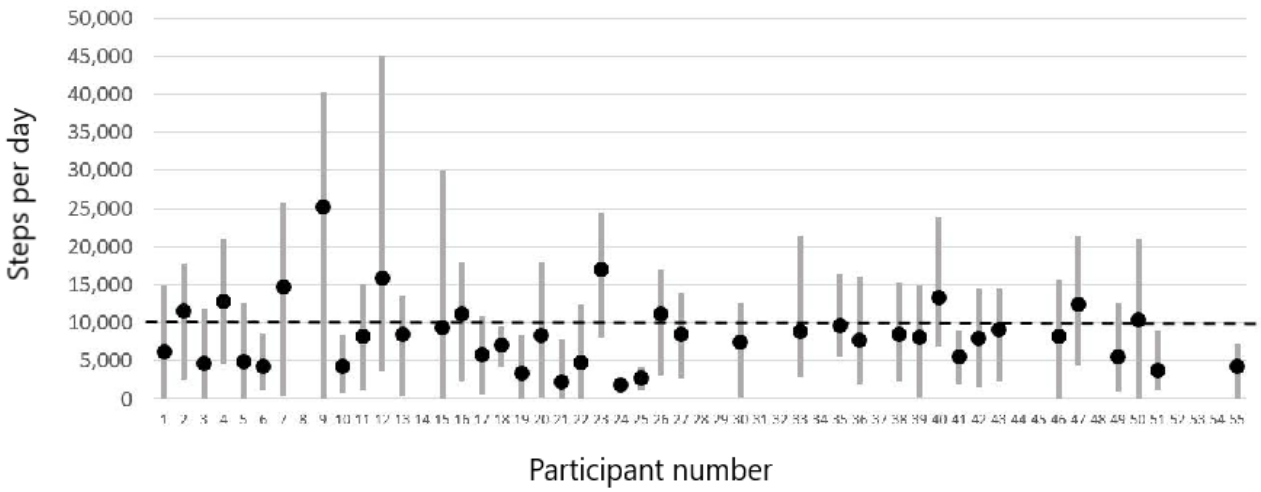
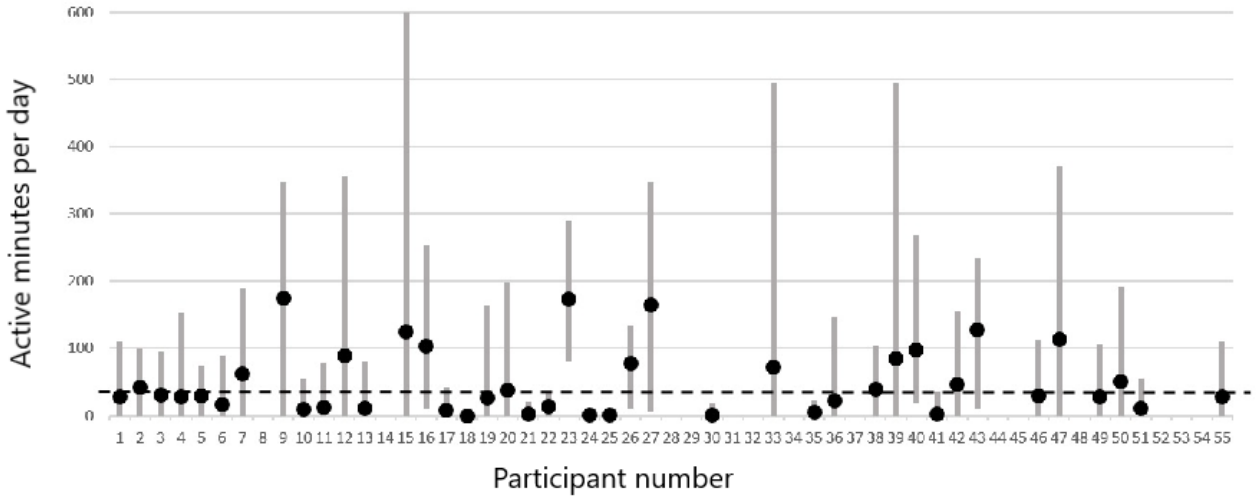


Figure 2 shows the second way of conceptualizing physical activity in this study: the number of active minutes per day. This measure may correspond more closely to the high level of physical activity that is needed for people with HIV to prevent cardiovascular complications of HIV infection. Similar to the steps-per-day metric, 59% (26/44) of the participants had cross-day averages above or close to the recommended level of 30 minutes per day. The active minutes measure had a higher *ICC*=0.78, meaning that active minutes were more consistently the same within individuals over time than the total minutes of activity. Only 39% of the variance was unexplained at the between-person level, but given the importance of high activity, and because of the fact that most people with averages <30 minutes per day still had some days with higher activity, it was also important to examine within-person predictors of active minutes.

Figure 2. Active minutes per day for each study participant. Black dots represent each person's average number of active minutes per day, across all days that they wore the Fitbit sensor device (up to 1 month for each participant). Gray bars represent the within-person average (SD 1). SDs were calculated within persons, and therefore, the varying height of the bars reflects the fact that some people's daily amount of physical activity varied more than that of others. The dashed line at 30 minutes per day reflects the recommended daily amount of active minutes for adults.



Intuitive-Level Predictors of Steps per Day

Table 1 shows the sensor and survey variables that predicted the number of steps taken by the same individual on the subsequent day. People with HIV took more steps the next day when they reported less fatigue, had higher sleep efficiency based on Fitbit sensor data, had higher percentages of time spent in deep sleep and rapid eye movement sleep, and woke up less often after going to bed. Having more steps was also related to lower stress based on higher HRV, although it was not statistically related to self-reported stress. Higher average HR, higher maximum HR, and lower minimum HR values also predicted the number of steps on the subsequent day, although these findings might simply reflect an association between more physical activity and better cardiovascular fitness. Among the survey variables, lower barriers to self-care, more social support, and more use of avoidance coping strategies each predicted the number of steps the following day. The finding on coping was in an unexpected direction and merits further investigation; however, all other relationships were consistent with the literature. In a follow-up exploratory analysis of individual items to determine the specific barriers that were the most important, effects were significant for alcohol use (β=−3311, $SE_β$ 875; P<.001), drug use (β=1153, $SE_β$ 256; P<.001), medication side effects (β=−6426, $SE_β$ 954; P<.001), feeling unwell (β=−1573, $SE_β$ 665; P=.02), feeling healthy (β=−2038, $SE_β$ 660; P=.002), and feeling irritated about having to take medications (β=−2428, $SE_β$ 621; P<.001), but not for forgetting, travel, or feeling confused.

In the multivariable model, the maximum HR and all of the sleep variables were no longer significant. As shown in the right-hand column of Table 1, the final model included average and minimum HR; higher HRV, indicating lower stress; lower self-reported fatigue based on PROMIS; higher self-reported social support; more use of avoidance coping; and fewer reported barriers to self-care.

Predictors of Active Minutes per Day

A smaller set of daily variables predicted high-intensity physical activity (Table 2). None of the sleep variables had significant effects in univariate analyses. Fewer HR metrics were significant, with higher maximum and average HR predicting active minutes the next day, but no relationship between high-intensity physical activity and minimum HR. Lower stress based on higher HRV predicted more active minutes the next day, and lower self-reported stress also predicted active minutes in a univariate model. In contrast to the findings about total steps, social support was nonsignificant as a predictor of active minutes, but HIV-related stigma (a related measure tied to negative social perceptions rather than positive ones) was significant and predicted active minutes in the opposite direction. Coping was significant in both analyses; however, high-intensity physical activity was predicted by the absence of approach coping rather than by the presence of avoidance coping. This is an inverse coping construct and, therefore, is in the same unexpected direction as the relationship between coping and total steps. The PROMIS fatigue scale and overall barriers to self-care scale again predicted active minutes. In the exploratory analyses, we found the same relationships between individual barriers and active minutes as were seen for total steps; the only exception was alcohol use, which predicted total steps but did not predict active minutes. Self-efficacy also predicted high-intensity physical activity, a relationship that was not seen in the analysis of total steps.

Maximum HR, self-efficacy, and stress were excluded from the multivariable model. Therefore, lower fatigue, less use of approach coping, less HIV-related stigma, and fewer perceived barriers were the only significant EMA predictors of next day active minutes, along with the sensor measures of average HR and lower HRV.

Table 1. Predictors of next day total steps.

Daily predictor variable	Univariate model			Combined model		
	β (SE)	*F* test (*df*)	*P* value	β (SE)	*F* test (*df*)	*P* value
HR[a] sensor data						
Resting HR (bpm[b])	4.74 (16.0)	0.09 (83)	.77	—[c]	—	—
Maximum HR (bpm)	33.8 (8.67)	15.2 (520)	<.001	—	—	—
Average HR (bpm)	133 (15.2)	76.3 (733)	<.001	325 (20.3)	256 (149)	.006
Minimum HR (bpm)	–45.2 (22.9)	3.88 (520)	.05	–215 (34.9)	38.2 (231)	.01
HRV[d] (mspb[e])	14.5 (4.33)	11.2 (52)	.008	22.8 (5.00)	20.8 (79)	.03
Sleep sensor data						
Total sleep time (minutes)	–1.43 (1.32)	1.18 (473)	.28	—	—	—
Time in bed (minutes)	–1.19 (1.16)	1.05 (467)	.31	—	—	—
Sleep efficiency (%)	–7663 (3658)	4.39 (376)	.04	—	—	—
Interrupted sleep (yes or no)	–719 (377)	3.65 (335)	.06	—	—	—
Wake after sleep onset (number)	96.9 (47.0)	4.25 (431)	.04	—	—	—
Light sleep (%)	–449 (1719)	0.07 (98)	.80	—	—	—
Deep sleep (%)	10,546 (2970)	12.6 (274)	.001	—	—	—
REM[f] sleep (%)	–6978 (1423)	24.0 (406)	.03	—	—	—
Daily survey data						
PROMIS[g] fatigue (*T* score)	–95.9 (9.80)	95.8 (338)	<.001	–46.6 (17.3)	7.25 (260)	.008
Self-efficacy (1-4 scale)	734 (435)	2.84 (41)	.09	—	—	—
Mood (1-4 scale)	66.2 (310)	0.05 (161)	.83	—	—	—
Stress (1-4 scale)	–414 (348)	1.41 (300)	.24	—	—	—
Avoidance coping (1-4 scale)	1114 (287)	15.1 (295)	.001	964 (244)	15.7 (245)	<.001
Approach coping (1-4 scale)	589 (336)	3.08 (429)	.08	—	—	—
Social support (1-4 scale)	822 (278)	871 (445)	.003	530 (254)	4.37 (237)	.04
Stigma (1-4 scale)	–29.1 (288)	0.01 (355)	.92	—	—	—
ART[h] adherence (yes or no)	–526 (1385)	0.14 (134)	.71	—	—	—
Motivation (1-7 scale)	97.3 (623)	0.02 (367)	.88	—	—	—
Barriers (1-4 scale)	–3590 (597)	36.1 (221)	<.001	–1396 (514)	7.37 (221)	.008

[a]HR: heart rate.

[b]bpm: beats per minute.

[c]Only statistically significant results are reported.

[d]HRV: heart rate variability.

[e]mspb: milliseconds per beat.

[f]REM: rapid eye movement.

[g]PROMIS: Patient-Reported Outcomes Measurement Information System.

[h]ART: antiretroviral treatment.

Table 2. Predictors of next day active minutes.

Daily predictor variable	Univariate model			Combined model		
	β (SE)	*F* test (*df*)	*P* value	β (SE)	*F* test (*df*)	*P* value
HR[a] sensor data						
Resting HR (bpm[b])	.26 (.26)	1.03 (528)	.31	—[c]	—	—
Maximum HR (bpm)	.91 (.10)	82.9 (99)	<.001	—	—	—
Average HR (bpm)	1.77 (.19)	90.9 (768)	<.001	3.21 (.26)	155 (284)	<.001
Minimum HR (bpm)	.35 (.26)	1.77 (512)	.18	—	—	—
HRV[d] (mspb[e])	.41 (.05)	55.1 (616)	<.001	.66 (.07)	79.6 (259)	<.001
Sleep sensor data						
Total sleep time (minutes)	–.02 (.02)	0.97 (488)	.33	—	—	—
Time in bed (minutes)	–.01 (.01)	0.67 (507)	.41	—	—	—
Sleep efficiency (%)	–41.9 (44.4)	0.89 (536)	.35	—	—	—
Interrupted sleep (yes or no)	–4.96 (4.67)	1.13 (517)	.29	—	—	—
Wake after sleep onset (number)	–.20 (.56)	0.13 (205)	.72	—	—	—
Light sleep (%)	–.01 (.02)	0.09 (98)	.76	—	—	—
Deep sleep (%)	–.03 (.04)	0.54 (274)	.46	—	—	—
REM[f] sleep (%)	–.04 (.06)	0.38 (406)	.54	—	—	—
Daily survey data						
PROMIS[g] fatigue (*T* score)	–1.81 (.27)	45.4 (185)	<.001	–.98 (0.25)	15.4 (287)	<.001
Self-efficacy (1-4 scale)	12.2 (5.30)	5.31 (72)	.02	—	—	—
Mood (1-4 scale)	7.43 (3.85)	3.72 (108)	.06	—	—	—
Stress (1-4 scale)	–9.45 (4.23)	4.99 (51)	.03	—	—	—
Avoidance coping (1-4 scale)	1.04 (3.69)	0.08 (165)	.78	—	—	—
Approach coping (1-4 scale)	–10.4 (4.22)	6.05 (198)	.02	–8.50 (3.77)	5.08 (390)	.03
Social support (1-4 scale)	6.07 (3.46)	3.09 (226)	.08	—	—	—
Stigma (1-4 scale)	–9.33 (3.47)	7.24 (50)	.01	–9.59 (3.40)	7.98 (355)	.005
ART[h] adherence (yes or no)	–2.27 (19.8)	0.01 (180)	.91	—	—	—
Motivation (1-7 scale)	.79 (8.39)	0.01 (307)	.93	—	—	—
Barriers (1-4 scale)	–35.5 (8.03)	19.6 (352)	<.001	–22.4 (7.33)	9.34 (386)	.002

[a]HR: heart rate.

[b]bpm: beats per minute.

[c]Only statistically significant results are reported.

[d]HRV: heart rate variability.

[e]mspb: milliseconds per beat.

[f]REM: rapid eye movement.

[g]PROMIS: Patient-Reported Outcomes Measurement Information System.

[h]ART: antiretroviral treatment.

Narrative-Level and Demographic Predictors of Physical Activity

Table 3 shows the effects of baseline measures on the 2 physical activity metrics, including demographic variables and self-report questionnaires. These self-report tools assessed some of the same constructs that were included in the daily surveys but did so by engaging the participant's narrative mind through questions about "average" or "typical" experiences over the past 7 days. Although such retrospective self-report tools are commonly used, prior research suggests that they are more strongly biased by memories, beliefs, and expectancies than in-the-moment questions asked via daily surveys [20].

Table 3. Effects of demographic and narrative-level predictors.

Baseline variable	Effect on total steps			Effect on active minutes		
	r	*t* test (*df*)	*P* value	*r*	*t* test (*df*)	*P* value
Demographic predictors						
Employed versus not	–0.19	–0.92 (22)	.36	–0.19	-0.93 (22)	.09
Gender (male or female)	0.003	0.01 (10)	.99	0.28	0.94 (10)	.37
Race (any minority status)	0.05	0.28 (26)	.78	0.04	0.19 (26)	.85
Age (years)	0.03	0.23 (54)	.82	0.18	1.42 (54)	.16
Baseline questionnaires						
Fatigue (PROMIS[a] tool)	0.02	0.15 (54)	.88	–0.15	-1.14 (54)	.26
Mood (PROMIS tool)	–0.28	–1.92 (54)	.06	–0.28	-2.22 (54)	.03[b]
Confusion (PROMIS tool)	0.24	1.92 (54)	.06	0.40	3.23 (54)	.002[c]
Sleep (PROMIS tool)	0.05	0.39 (54)	.70	0.05	0.31 (54)	.76
Pain (PROMIS tool)	–0.08	–0.61 (54)	.54	–0.03	-0.23 (54)	.82
Stress (HIV-QoL[d] subscale)	–0.25	–1.97 (54)	.06	–0.11	-0.81 (54)	.42
Stigma (HIV Stigma scale)	–0.12	–0.92 (54)	.36	–0.32	-2.66 (54)	.01[b]

[a]PROMIS: Patient-Reported Outcomes Measurement Information System.

[b]Significant at $P<.05$.

[c]Significant at $P<.01$.

[d]HIV-QoL: HIV-related Quality of Life scale.

None of the demographic variables predicted either of the physical activity variables, and none of the baseline questionnaire measures had any effect on the total steps measure of physical activity. Three survey measures predicted active minutes: mood, cognitive confusion, and stigma. Some variables that were significant as day-by-day measures at the intuitive level of analysis failed to predict physical activity when measured at the narrative level, including stress and fatigue. Mood and cognitive confusion had the opposite pattern, with small to moderate effects on active minutes when measured at the narrative level only. Stigma predicted active minutes when measured at either the narrative or intuitive level, although it had a stronger relationship with active minutes when measured day by day.

Discussion

Principal Findings

The daily physical activity of people with HIV was predicted based on their previous day environments, experiences, and behaviors. Overall, this study found substantial within-person variability in physical activity and illustrated the importance of everyday experiences and behaviors in understanding daily levels of activity among people with HIV, as suggested by TMT. To our knowledge, this is the first study to examine EMA daily survey measures as predictors of physical activity in people with HIV. Survey-based predictors of overall physical activity in people with HIV included greater avoidance coping, higher perceived social support, and fewer reported barriers to self-care, including the absence of alcohol use, drug use, ART side effects, feeling unwell, and irritation about medication use. The finding on coping was in an unexpected direction; however, other findings were consistent with barriers to physical activity that have been reported in previous studies of people with HIV [13,14].

This study also examined sensor-based predictors of physical activity, which provided data on everyday experiences that might have affected the behavior of people with HIV outside of their conscious awareness. Significant predictors of next day total steps included maximum and average HR, both of which might reflect cardiovascular fitness, as well as HRV, which is a physiological indicator of stress. In contrast to this physiologically based stress metric, participants' self-reported stress was a less useful predictor that had no association with next day total steps, although it weakly predicted next day active minutes in a univariate model. On the other hand, next day physical activity was more consistently related to subjective fatigue based on the PROMIS tool than to sensor-based measures of sleep quality.

Findings about high-intensity physical activity were generally consistent with those for overall activity, although there were fewer significant predictors for the high-intensity level of physical activity that is more important in preventing cardiovascular disease. High fatigue, high levels of HIV-related stigma, and high stress based on HRV are risk factors for inactivity that can be identified by monitoring everyday psychological states. All of these are potentially modifiable risk factors, which could be addressed by clinicians either directly or with a referral to mental health support services. In a prior study, poor mood was a unique prospective marker for days when people with HIV were less likely to miss a dose of ART

medication [19]. Although the mood scale in this study did not significantly predict exercise adherence, related constructs such as stress and fatigue showed significant prospective relationships with next day physical activity.

Implications for TMT

In this study, variables from EMA surveys and wrist-worn sensors were found to prospectively predict participants' physical activity on the day after the variable was measured. The daily experiences measured in this study are considered to represent the intuitive system identified by TMT [17], as their closeness in time to the actual point of behavior reduces the chance of retrospective mental editing of experiences. Furthermore, some of the data gathered via the sensors were immune to self-reporting bias and might not even have been accessible using standard retrospective questionnaire methods. Daily within-person variations in behavior accounted for 51% of the total variability in steps per day and 39% of the variability in active minutes, illustrating the importance of testing day-by-day predictors of physical activity. The time-lagged analysis in this study also allowed us to draw prospective conclusions that are stronger than purely correlational findings. Although our results cannot prove a causal relationship because nothing was manipulated, they do show a cause preceding an effect, a feature often lacking in correlational studies.

For a direct comparison between narrative- and intuitive-level predictors, we also examined the effects of stable participant demographic characteristics and the effects of predictor variables assessed using standard retrospective questionnaires at a single point in time. None of these variables were significantly associated with steps per day, and only three of the questionnaire measures—depressed mood, cognitive confusion, and HIV stigma—significantly predicted active minutes. The best narrative-level measures had small to moderate effects, despite the fact that we used best practice PROMIS tools to measure constructs of interest. In general, our findings suggest that intuitive-level measures are more useful than narrative-level measures for predicting the physical activity of people with HIV, although one of the predictors (mood) was significant only when it was measured at the narrative level.

Strengths and Limitations

Participants in this study were people with HIV recruited from a Ryan White clinic in Denver, Colorado, United States. Of the 1.2 million people currently living with HIV in the United States, approximately 59% are currently in care [37], and of those in care, >80% receive their medical treatment through Ryan White clinics [38]. In that sense, the participants in this study were representative of people with HIV receiving care in the United States. However, our sample was disproportionately White, reflecting the demographics of Colorado, whereas the US population of people with HIV includes higher percentages of people who are Black and Latinx. Therefore, our findings may not be applicable to people of color with HIV.

This study was also limited by a moderate sample size, which may reduce generalizability, although the collection of data on multiple days from each participant increased the effective sample size for multilevel analyses. We used a minimally restrictive α level of .05 to identify potential predictors of physical activity in this initial study of everyday experiences; however, there is a risk of type 1 error in our findings. Future studies with larger sample sizes might use more restrictive statistical assumptions to further narrow the number of predictors of physical activity. In addition, Fitbits are not generally viewed as research-grade devices, although they have shown an ability to differentiate sedentary individuals from moderately active individuals in prior studies. Future research could further clarify the relative importance of predictor variables on physical activity by using research-grade activity sensors that have less measurement error, thereby providing greater statistical power for analyses.

Finally, there may be some measurement challenges in this study: the interpretation of HR metrics, including HRV, is not always clear and may be affected by overall cardiovascular fitness and participants' psychological states. Other daily measures were self-report surveys, and although the tools were validated in prior EMA studies, there is some potential for bias based on self-presentation or inaccurate recall. Findings on coping were in an unexpected direction, which might also be an indication of measurement problems specific to this construct; for example, a participant's use of more coping strategies might suggest ineffective or inefficient coping, whereas a high score on just 1 of the 9 items might suggest successful coping. Alternatively, it might be that some participants specifically used exercise as a form of coping; although exercise is a healthy behavior, it could be considered avoidance coping as it does not directly engage with the source of stress. If this is the explanation, it might even be the case that exercising close in time to the source of stress results in more effective coping later on (eg, at a time point that is 2 steps removed from the initial stressor). Thus, the unexpected finding about coping presents intriguing possibilities that could benefit from future research.

Implications for Practice and Research

Our findings illustrate the need for improved physical activity among people with HIV based on day-to-day variability, even among people with HIV who had average daily activity levels above those recommended by health promotion guidelines. Within-person variability was particularly apparent with regard to high-intensity physical activity. Given that everyday experiences such as stress, fatigue, and HIV-related stigma were found to interfere with high-intensity physical activity, clinicians should ask people with HIV about their stress and fatigue. Clinicians can also help their patients develop strategies to address barriers such as stigma and provide referrals to mental health resources, as appropriate. Finally, clinical care environments should take steps to reduce HIV-related stigma and make people with HIV feel welcomed. This may help to reduce the overall barriers to health care for people with HIV, as well as support their physical activity.

Further research is needed on interventions to increase physical activity among people with HIV. In-the-moment interventions might target improvement in EMA-measured variables such as HIV-related stigma and fatigue, as well as sensor-based predictors such as stress based on HRV. TMT suggests that different variables might be important in initiating versus

maintaining physical activity over time; however, this study did not differentiate new versus established exercise habits, which might be another important question to investigate in future intervention studies. Finally, research on physical activity using EMA is still relatively novel, and questions related to the accuracy of measurements using both sensors and surveys are also important topics for ongoing investigation. Future studies using both EMA and sensor data can enhance our understanding of the physical activity and other everyday behaviors of people with HIV.

Conclusions

Overall, this study suggests an opportunity to improve how people with HIV manage everyday challenges in order to enhance their physical activity. People with HIV reported significant levels of fatigue, which predicted both total steps and high-intensity physical activity. Subjective fatigue was a better predictor of high-intensity physical activity than actual sleep stages, as estimated using the Fitbit device. In addition, stress predicted physical activity; however, a physiological stress measure, HRV, was a stronger predictor of both total steps and active minutes than self-reported daily stress. Other daily psychological experiences, including self-efficacy, coping, and HIV-related stigma, also predicted physical activity, as did everyday self-care barriers such as alcohol use and ART side effects. Many of these are potentially modifiable variables that could be targeted by clinicians or in future research interventions to improve the physical activity of people with HIV.

Acknowledgments

This study received funding from the University of Colorado Nursing Biobehavioral Area of Excellence and support from the Colorado Clinical and Translational Science Institute, National Institutes of Health UL1 TR002535.

Conflicts of Interest

None declared.

References

1. Fauci A, Redfield R, Sigounas G, Weahkee MD, Giroir BP. Ending the HIV epidemic: a plan for the United States. JAMA 2019 Mar 05;321(9):844-845 [FREE Full text] [doi: 10.1001/jama.2019.1343] [Medline: 30730529]
2. Pathai S, Bajillan H, Landay AL, High KP. Is HIV a model of accelerated or accentuated aging? J Gerontol A Biol Sci Med Sci 2014 Jul;69(7):833-842 [FREE Full text] [doi: 10.1093/gerona/glt168] [Medline: 24158766]
3. Barroso J, Voss J. Fatigue in HIV and AIDS: an analysis of evidence. J Assoc Nurses AIDS Care 2013;24(1 Suppl):S5-14 [FREE Full text] [doi: 10.1016/j.jana.2012.07.003] [Medline: 23290377]
4. Burke E, Nelson J, Kwong J, Cook P. Cardiovascular risk assessment for persons living with HIV. J Assoc Nurses AIDS Care 2012;23(2):134-145 [FREE Full text] [doi: 10.1016/j.jana.2011.05.007] [Medline: 21775164]
5. Reddy K, Parker R, Losina E, Baggett TP, Paltiel AD, Rigotti NA, et al. Impact of cigarette smoking and smoking cessation on life expectancy among people with HIV: a US-based modeling study. J Infect Dis 2016 Dec 01;214(11):1672-1681 [FREE Full text] [doi: 10.1093/infdis/jiw430] [Medline: 27815384]
6. Webel A, Barkley J, Longenecker C, Mittelsteadt A, Gripshover B, Salata R. A cross-sectional description of age and gender differences in exercise patterns in adults living with HIV. J Assoc Nurses AIDS Care 2015;26(2):176-186 [FREE Full text] [doi: 10.1016/j.jana.2014.06.004] [Medline: 25249267]
7. Erlandson K, Schrack J, Jankowski C, Brown T, Campbell T. Functional impairment, disability, and frailty in adults aging with HIV-infection. Curr HIV/AIDS Rep 2014 Sep;11(3):279-290 [FREE Full text] [doi: 10.1007/s11904-014-0215-y] [Medline: 24966138]
8. Erlandson K, MaWhinney S, Wilson M, Gross L, McCandless SA, Campbell TB, et al. Physical function improvements with moderate or high-intensity exercise among older adults with or without HIV infection. AIDS 2018 Oct 23;32(16):2317-2326 [FREE Full text] [doi: 10.1097/QAD.0000000000001984] [Medline: 30134299]
9. Sturm R, Cohen D. Free time and physical activity among Americans 15 years or older: cross-sectional analysis of the American time use survey. Prev Chronic Dis 2019 Sep 26;16:E133 [FREE Full text] [doi: 10.5888/pcd16.190017] [Medline: 31560643]
10. Nosrat S, Whitworth J, Ciccolo J. Exercise and mental health of people living with HIV: a systematic review. Chronic Illn 2017 Dec;13(4):299-319 [FREE Full text] [doi: 10.1177/1742395317694224] [Medline: 29119865]
11. Montoya J, Jankowski C, O'Brien K, Webel AR, Oursler KK, Henry BL, et al. Evidence-informed practical recommendations for increasing physical activity among persons living with HIV. AIDS 2019 May 01;33(6):931-939 [FREE Full text] [doi: 10.1097/QAD.0000000000002137] [Medline: 30946147]
12. Bhochhibhoya A, Qiao S, Mao Y, Zhou Y, Shen Z, Sui X, et al. Physical activity, depression, and antiretroviral therapy adherence among people living with HIV: a mediation analysis. J Assoc Nurses AIDS Care 2019;30(5):556-562. [doi: 10.1097/JNC.0000000000000045] [Medline: 30633084]
13. Vancampfort D, Mugisha J, Richards J, De Hert M, Probst M, Stubbs B. Physical activity correlates in people living with HIV/AIDS: a systematic review of 45 studies. Disabil Rehabil 2018 Jul;40(14):1618-1629 [FREE Full text] [doi: 10.1080/09638288.2017.1306587] [Medline: 28325087]

14. Quigley A, Baxter L, Keeler L, MacKay-Lyons M. Using the Theoretical Domains Framework to identify barriers and facilitators to exercise among older adults living with HIV. AIDS Care 2019 Feb;31(2):163-168 [FREE Full text] [doi: 10.1080/09540121.2018.1499860] [Medline: 30021454]
15. Prochaska J, Velicer W, Rossi J, Goldstein M, Marcus B, Rakowski W, et al. Stages of change and decisional balance for 12 problem behaviors. Health Psychol 1994;13(1):39-46 [FREE Full text] [doi: 10.1037/0278-6133.13.1.39]
16. Tulloch H, Sweet S, Fortier M, Capstick G, Kenny G, Sigal R, et al. Exercise facilitators and barriers from adoption to maintenance in the diabetes aerobic and resistance exercise trial. Can J Diabetes 2013 Dec;37(6):367-374 [FREE Full text] [doi: 10.1016/j.jcjd.2013.09.002] [Medline: 24321716]
17. Cook P, Schmiege S, Reeder B, Horton-Deutsch S, Lowe N, Meek P. Temporal immediacy: a two-system theory of mind for understanding and changing health behaviors. Nurs Res 2018;67(2):108-121 [FREE Full text] [doi: 10.1097/NNR.0000000000000265] [Medline: 29489632]
18. Kahneman D. Thinking, Fast and Slow. New York: Penguin Books; 2013.
19. Cook P, Schmiege S, Starr W, Carrington J, Bradley-Springer L. Prospective state and trait predictors of daily medication adherence behavior in HIV. Nurs Res 2017;66(4):275-285 [FREE Full text] [doi: 10.1097/NNR.0000000000000216] [Medline: 28654566]
20. Shiffman S, Stone A, Hufford M. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32 [FREE Full text] [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
21. Reeder B, Cook P, Meek P, Ozkaynak M. Smart watch potential to support augmented cognition for health-related decision making. In: Augmented Cognition. Neurocognition and Machine Learning. Cham: Springer; 2017.
22. Makic M, Gilbert D, Jankowski C, Reeder B, Al-Salmi N, Starr W, et al. Sensor and survey measures associated with daily fatigue in HIV: findings from a mixed-method study. J Assoc Nurses AIDS Care 2020;31(1):12-24 [FREE Full text] [doi: 10.1097/jnc.0000000000000152]
23. Cook P, Hartson K, Schmiege S, Jankowski C, Starr W, Meek P. Bidirectional relationships between fatigue and everyday experiences in persons living with HIV. Res Nurs Health 2016 Jun;39(3):154-163 [FREE Full text] [doi: 10.1002/nur.21718] [Medline: 27059443]
24. Patient-Reported Outcomes Measurement Information System (PROMIS). U.S. Department of Health and Human Services. URL: https://www.nia.nih.gov/research/resource/patient-reported-outcomes-measurement-information-system-promis#:~:text=The%20Patient%2DReported%20Outcomes%20Measurement,health%20in%20adults%20and%20children [accessed 2022-03-10]
25. Holmes W, Shea J. A new HIV/AIDS-targeted quality of life (HAT-QoL) instrument: development, reliability, and validity. Med Care 1998 Feb;36(2):138-154 [FREE Full text] [doi: 10.1097/00005650-199802000-00004] [Medline: 9475469]
26. Berger B, Ferrans C, Lashley F. Measuring stigma in people with HIV: psychometric assessment of the HIV stigma scale. Res Nurs Health 2001 Dec;24(6):518-529 [FREE Full text] [doi: 10.1002/nur.10011] [Medline: 11746080]
27. Dobbs W, Fedewa M, MacDonald H, Holmes CJ, Cicone ZS, Plews DJ, et al. The accuracy of acquiring heart rate variability from portable devices: a systematic review and meta-analysis. Sports Med 2019 Mar;49(3):417-435 [FREE Full text] [doi: 10.1007/s40279-019-01061-5] [Medline: 30706234]
28. Evenson K, Goto M, Furberg R. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12:159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
29. What are Active Zone Minutes or active minutes on my Fitbit device? Fitbit, LLC. URL: https://help.fitbit.com/articles/en_US/Help_article/1379.htm [accessed 2022-03-10]
30. Jetté M, Sidney K, Blümchen G. Metabolic equivalents (METS) in exercise testing, exercise prescription, and evaluation of functional capacity. Clin Cardiol 1990 Aug;13(8):555-565 [FREE Full text] [doi: 10.1002/clc.4960130809] [Medline: 2204507]
31. Pulse watch. GitHub. URL: https://iccir919.github.io/pulseWatch/public/index.html [accessed 2022-03-10]
32. Kamarck T. The diary of ambulatory behavioral states: a new approach to the assessment of psychosocial influences on ambulatory cardiovascular activity. In: Technology and Methods in Behavioral Medicine. East Sussex, United Kingdom: Psychology Press; 1998.
33. Stone A, Neale J. New measure of daily coping: development and preliminary results. J Personality Soc Psychol 1984;46(4):892-906 [FREE Full text] [doi: 10.1037/0022-3514.46.4.892]
34. Herzog T, Blagg C. Are most precontemplators contemplating smoking cessation? Assessing the validity of the stages of change. Health Psychol 2007 Mar;26(2):222-231 [FREE Full text] [doi: 10.1037/0278-6133.26.2.222] [Medline: 17385974]
35. Chesney M, Ickovics J, Chambers D, Gifford AL, Neidig J, Zwickl B, et al. AIDS Care 2000 Jun;12(3):255-266. [doi: 10.1080/09540120050042891] [Medline: 10928201]
36. Luke D. Multilevel Modeling. Thousand Oaks, California: SAGE Publications; 2004.
37. Sullivan P, Woodyatt C, Koski C, Pembleton E, McGuinness P, Taussig J, et al. A data visualization and dissemination resource to support HIV prevention and care at the local level: analysis and uses of the AIDSVu public data resource. J Med Internet Res 2020 Oct 23;22(10):e23173 [FREE Full text] [doi: 10.2196/23173] [Medline: 33095177]

38. Ryan White HIV/AIDS program annual client-level data report. U.S. Health Resources and Services Administration. URL: https://hab.hrsa.gov/sites/default/files/hab/data/datareports/RWHAP-annual-client-level-data-report-2019.pdf [accessed 2022-03-10]

Abbreviations

ART: antiretroviral treatment
DABS: Diary of Ambulatory Behavioral States
EMA: ecological momentary assessment
HR: heart rate
HRV: heart rate variability
ICC: intraclass correlation coefficient
PROMIS: Patient-Reported Outcomes Measurement Information System
REDCap: Research Electronic Data Capture
TMT: Two Minds Theory

Edited by L Buis; submitted 04.10.21; peer-reviewed by J Voss, Y Cai, R Marshall; comments to author 06.12.21; revised version received 13.01.22; accepted 17.02.22; published 14.04.22.

Please cite as:
Cook P, Jankowski C, Erlandson KM, Reeder B, Starr W, Flynn Makic MB
Low- and High-Intensity Physical Activity Among People with HIV: Multilevel Modeling Analysis Using Sensor- and Survey-Based Predictors
JMIR Mhealth Uhealth 2022;10(4):e33938
URL: https://mhealth.jmir.org/2022/4/e33938
doi:10.2196/33938
PMID:35436236

Original Paper

Chronic Tinnitus and the Positive Effects of Sound Treatment via a Smartphone App: Mixed-Design Study

Justyna Jolanta Kutyba[1], MSc; W Wiktor Jędrzejczak[1], Prof Dr; Elżbieta Gos[1], PhD; Danuta Raj-Koziak[1], Prof Dr; Piotr Henryk Skarzynski[1,2,3], Prof Dr

[1]World Hearing Center, Institute of Physiology and Pathology of Hearing, Warsaw, Poland

[2]Institute of Sensory Organs, Kajetany, Poland

[3]Department of Heart Failure and Cardiac Rehabilitation, Second Faculty, Medical University of Warsaw, Warsaw, Poland

Corresponding Author:
Justyna Jolanta Kutyba, MSc
World Hearing Center
Institute of Physiology and Pathology of Hearing
Mokra 17
Kajetany
Warsaw, 05-830
Poland
Phone: 48 222769572
Email: j.kutyba@ifps.org.pl

Abstract

Background: Tinnitus is a phantom auditory sensation in the absence of an external stimulus. It is accompanied by a broad range of negative emotional symptoms and a significantly lower quality of life. So far, there is no cure for tinnitus, although various treatment options have been tried. One of them is mobile technology employing dedicated apps based on sound therapy. The apps can be managed by the patient and tailored according to their needs.

Objective: The study aims to assess the effect of a mobile app that generates background sounds on the severity of tinnitus.

Methods: The study involved 68 adults who had chronic tinnitus. Participants were divided into a study group (44 patients) and a control group (24 patients). For 6 months those in the study group used a free mobile app that enriched the sound environment with a background sound. Participants were instructed to use the app for at least 30 minutes a day using their preferred sound. The participants in the control group did not use the app. Subjective changes in the day-to-day functioning of both groups were evaluated using the Tinnitus Handicap Inventory (THI) questionnaire, a visual analog scale, and a user survey.

Results: After 3 months of using the app, the THI global score significantly decreased (P<.001) in the study group, decreasing again at 6 months (P<.001). The largest improvements were observed in the emotional and catastrophic reactions subscales. A clinically important change in the THI was reported by 39% of the study group (17/44). Almost 90% of the study participants (39/44) chose environmental sounds to listen to, the most popular being rain and ocean waves. In the control group, tinnitus severity did not change over 3 or 6 months.

Conclusions: Although the participants still experienced limitations caused by tinnitus, the advantage of the app was that it led to lower negative emotions and thus reduced overall tinnitus severity. It is worth considering whether a mobile app might be incorporated into the management of tinnitus in a professional setting.

(JMIR Mhealth Uhealth 2022;10(4):e33543) doi:10.2196/33543

KEYWORDS

tinnitus; mobile app; smartphone; sound therapy; telemedicine; mobile phone

Introduction

The digital revolution initiated in the 20th century has impacted many fields of life, including those related to health [1]. The use of technological solutions to improve treatments can be seen in every branch of medicine. Examples are solutions with social benefits, such as e-referrals and e-prescriptions [2], as well as those with individual benefits such as health monitoring wristbands [3] and virtual assistants [4]. Modern solutions are

evident in cardiology [5], diabetology [6], and audiology [7], including patients with tinnitus.

Tinnitus is a sound that occurs without physical external stimulation and in most cases is heard only subjectively [8,9]. Tinnitus can be described as a squeaking, buzzing, humming, or clicking, either in the ears or in the head [8,10]. Epidemiology suggests the problem affects between 4% and 15% of the adult population and shows an increasing trend [10,11].

Currently, there is no established effective treatment. Experts dealing with tinnitus most often recommend sound therapy, which is the use of additional background sounds to change the patient's reaction to tinnitus [10]. Sound therapy does not cure the condition, but may significantly lower its severity, reducing the level of distress or impact that tinnitus has on the individual [10]. The chronic nature of tinnitus often prompts patients to use various forms of self-help [10,12]. Studies have shown that self-help programs can be effective and provide benefits to patients in managing their tinnitus [13-15]. It has also been documented that one of the most commonly used forms of self-help is acoustic stimulation, that is, surrounding oneself with various types of sounds [13,16]. Sound enrichment causes the patient's tinnitus to blend in with the environment and become less noticeable, thus improving overall comfort [17,18]. To enrich the auditory background, patients can use professional devices such as sound generators or broadband noise generators, as well as recorded CDs, the radio, and recently, mobile app [10,19].

In the mobile solutions market, a range of apps have been developed for people with tinnitus [20-22]: tutorials, apps containing information about tinnitus, apps to support psychotherapy, and others to gauge tinnitus characteristics or track tinnitus intensity throughout the day [23]. Studies show that the most commonly used apps generate sounds of nature, of everyday life, or relaxing music [23]. Patients use the apps because they are mostly free and easily available, and their variety makes it easy to select one that suits the individual.

The use of mobile apps in the treatment of tinnitus has been of interest to researchers for several years. There have been some papers describing the effects that mobile apps which generate sounds have on tinnitus. Henry et al [24] conducted a study on a group of 25 individuals who used a mobile sound-generating app for 8 weeks. The changes in tinnitus sensation as assessed by the Tinnitus Functional Index (TFI) questionnaire were generally small, although 8 of the 25 individuals achieved a reduction in total score of 13 points or more. This degree of reduction reflects a change that is likely to be clinically important to the patient.

Another study of app use was conducted in 2016 by Tyler et al [25] in a group of 16 individuals with a cochlear implant. The study assessed the level of acceptance of the sounds generated from the app. The participants reported that using an app was satisfactory and more comfortable than using tablets or computers. The participants indicated that the app provided them with a wide range of sounds and that the sounds provided to the implant were acceptable and pleasant.

In 2020 the current authors conducted a pilot study on 52 patients with tinnitus [26]. Participants listened to the mobile app daily for 6 months. Two questionnaires (TFI and Tinnitus Handicap Inventory [THI]) were used to evaluate the effect of the app on tinnitus severity. As gauged by both questionnaires, a significant reduction in tinnitus severity was found. Overall severity decreased after the first 3 months and again over the subsequent 3 months.

Although all these results are promising, these studies require further substantiation because only small groups of patients were studied, standardized tinnitus instruments were not always used, and there were no control groups. Thus, this study aimed to assess the effect of using a smartphone app that generates acoustic signals on the severity of tinnitus. The effect was evaluated by validated tinnitus-specific questionnaires after 3 and 6 months of use, and the results were compared with those obtained from people with tinnitus who did not use the app.

Methods

Participants and Setting

The participants were patients admitted to our tertiary referral center due to tinnitus. Only adult patients who had tinnitus for at least 6 months were recruited. The study's purpose and procedures were explained to all potential participants. Afterward, those who were willing to participate and owned an Android or iOS smartphone were included in the study group; those who did not own an appropriate device or did not want to participate in the study group but wanted to have their tinnitus monitored were included in the control group.

A total of 147 participants were initially enrolled in the study, but the final analysis included results from 68 participants, made up of 44 in the study group and 24 in the control group. There were 79 participants who were excluded, made up of 50 patients initially recruited to the study group and 29 patients recruited to the control group. They were excluded due to the following reasons: (1) 52 (30 enrolled in the study group and 22 in the control group) did not send back the questionnaires; (2) 17 (14 enrolled in the study group and 3 in the control group) sent back questionnaires with incomplete responses and were unsuitable for further analysis; (3) 10 (6 enrolled in the study group and 4 in the control group) admitted they had used other forms of therapy while participating in the study.

The dropout rate was similar in both groups: 53% (50/94) in the study group and 55% (29/53) in the control group. In the study group the dropouts were of similar age (mean 52.7 [SD 13.6]) as those who completed the study (mean 51.9 [SD 11.6]; t_{92}=0.30, P=.77). Tinnitus severity as measured with the THI global score was similar in the dropouts (mean 54.9 [SD 26.6]) and in the completers (mean 54.9 [SD 23.9]; t_{92}=0.02; P=.99). Their results on all 3 THI subscales were also similar in the study group.

The same was true for the control group. The dropouts were of similar age (mean 51.6 [SD 13.0]) as those who completed the study (mean 51.9 [SD 14.1]; t_{51}=0.09; P=.93). Tinnitus severity as measured with the THI global score was similar in the

dropouts (mean 43.7 [SD 22.3]) and in the completers (mean 52.1 [SD 23.1]; t_{51}=0.1.32; *P*=.19). Besides, their results on all 3 THI subscales were similar in the control group.

Patients in the study and control groups did not differ significantly in terms of sociodemographic data (Table 1).

Table 1. Characteristics of the participants.

Characteristics	Study group (n=44)	Control group (n=24)	Test result
Age (years)			t_{66}=0.01; *P*=.99
Mean (SD)	51.9 (11.6)	51.9 (14.1)	
Range	26-72	28-74	
Sex, n (%)			χ_1^2=0.04; *P*=.83
Female	25 (57)	13 (54)	
Male	19 (43)	11 (46)	
Educational status, n (%)[a]			χ_1^2=0.14; *P*=.71
No higher education	19 (45)	12 (50)	
Higher education	23 (55)	12 (50)	
Place of residence, n (%)			χ_1^2=0.68; *P*=.41
Rural	9 (20)	3 (13)	
Urban	35 (80)	21 (87)	
Tinnitus duration (years)			t_{66}=0.82; *P*=.42
Mean (SD)	6.2 (8.1)	4.7 (5.4)	
Range	0.7-40	1-26	
Tinnitus localization, n (%)[b]			χ_1^2=2.57; *P*=.11
One ear	13 (30)	12 (50)	
Both ears	30 (70)	12 (50)	
Manifestation over time, n (%)[b]			χ_1^2=0.04; *P*=.84
Constant	40 (93)	22 (92)	
Intermittent	3 (7)	2 (8)	
Hearing loss, n (%)			χ_1^2=0.11; *P*=.74
Yes	22 (50)	13 (54)	
No	22 (50)	11 (46)	

[a]n=42 for the study group because 2 participants did not state this.

[b]n=43 for the study group because 1 participant did not state this.

ReSound Tinnitus Relief App

The ReSound Tinnitus Relief app was tested in our pilot study, and because of promising results, it was used again in this study [26,27]. The app was created in 2014 and is available on the Google Play Store and the App Store. Before starting the study, we tested the operation of the app on both platforms. As the operation of the app on both platforms was the same, in our study we did not control which operating system the participants used. This particular app was chosen because it is free, has a menu in Polish, and has already been described in the literature [24,25,27,28]. The main function of the app is to generate various acoustic signals. It is equipped with a library of 33 sounds divided into 3 categories: environment, music, and therapeutic sounds. In addition, it offers meditation support, a set of relaxation exercises, and a panel with information about tinnitus. Participants included in the study group were given the same recommendations for using the app as those in the pilot study [26]: duration of stimulation should be a minimum of 30 minutes per day and the sound should be emitted in free field and at a level slightly lower than the patient's tinnitus. Patients decided for themselves which sounds they would listen to [26]. The patients in the control group did not use the app.

Measures

To assess the patients, the Tinnitus Handicap Inventory, a visual analog scale (VAS), and a survey were used.

The THI is a standardized questionnaire to assess the impact of tinnitus on a patient's daily functioning [29]. It is composed of 25 questions divided into 3 subscales: Functional, Emotional,

and Catastrophic. For each question, the patient can answer *yes*, *sometimes*, or *no*. The maximum score is 100 points; the higher the score, the more severe the tinnitus. A reduction in total score of 20 or more points is considered a clinically significant change in tinnitus severity [30]. The Polish version of the questionnaire was used in this study; its psychometric values have been determined by Skarżyński et al [31,32].

VAS is a tool that allows the assessment of a patient's tinnitus [33]. It presents a 10-cm-long line printed on a sheet of paper, with the ends of the line labeled "minimally" (at the left) and "maximally" (at the right). To calculate a VAS score, the length of the segment from the beginning of the line to the point indicated by the patient is measured and multiplied by 10, giving a number between 0 and 100. In this way, patients are asked to assess their level of tinnitus loudness and tinnitus annoyance. The VAS has been used in other studies of tinnitus [34-36].

A user survey was developed to assess how the participants used the app. The questions concerned the amount of time of using the app, the type of sounds listened to, subjective assessment of its effectiveness, and the use of other concurrent forms of therapy. The controls were given a modified version of the survey without questions relating to the app. The questions asked were: *How often do you use the application? How long do you use the application during the day? Have you made any other efforts at therapy while using the application?* Participants who answered "yes" to using other forms of therapy (n=10) were not included in the final analysis.

Patient Assessment Procedure

All patients signed an informed consent form to participate in the study. Afterward, a medical examination and pure tone audiometry were conducted. The app was installed on the patient's smartphone, information on how to use it was provided, and the patients filled in the THI and VAS questionnaires. Then, 3 and 6 months later, the participants received another set of questionnaires in the postal mail consisting of the THI, VAS, and the survey, and were again asked to fill them in and send them back.

Statistical Analysis

The demographic and clinical characteristics of the study and control groups were examined using descriptive statistics and percentages. Differences across groups were assessed through an independent (unpaired) *t* test or a χ^2 test. A mixed-design ANOVA with Bonferroni adjustment for multiple comparisons was used to evaluate tinnitus severity, loudness, and annoyance in the study and control groups over 6 months. A *P* value <.05 was considered statistically significant. The analysis was conducted using IBM SPSS Statistics (version 24).

Ethics Approval

The study protocol was based on a previously conducted pilot study [26] and was approved by the Bioethics Committee of the Institute of Hearing Physiology and Pathology (KB. IFPS:15/2020).

Results

Tinnitus Severity as Measured With the THI

Descriptive statistics for the THI scores for initial and follow-up measurements in the study and control groups are shown in Tables 2 and 3.

A mixed-design ANOVA revealed that an interaction effect (group × time) was statistically significant for the THI global score and for 2 of the 3 subscales (Emotional and Catastrophic). The interaction effect for the THI global score was $F_{2,132}$=6.15; *P*=.003; e^2=0.085. After 3 months of using ReSound the THI global score significantly decreased in the study group (*P*<.001), and at the 6-month follow-up it was significantly lower in comparison with the initial level (*P*<.001). However, there was no statistically significant difference (*P*=.23) between the 3- and 6-month follow-ups. In the control group, scores were similar at the initial measurement and at both follow-ups.

The same was true for the Emotional subscale, where the interaction effect was statistically significant ($F_{2,132}$=9.79; *P*<.001; e^2=0.139). Again, 3 months of using ReSound resulted in a significant decrease of tinnitus severity in the study group (*P*<.001), and at the 6-month follow-up tinnitus severity was significantly lower in comparison with the initial level (*P*<.001). There was no statistically significant difference (*P*=.13) between the 3- and 6-month follow-ups. Tinnitus severity remained stable in the control group across all analyzed measurements.

The interaction effect for the Catastrophic subscale was statistically significant ($F_{2,132}$=11.97; *P*<.001; e^2=0.185). Multiple comparisons showed that tinnitus severity significantly decreased in the study group at the 3-month follow-up (*P*<.001) and again significantly decreased after another 3 months (*P*=.01). The scores on the Catastrophic subscale remained stable in the control group.

For the Functional subscale the interaction effect was statistically nonsignificant ($F_{2,132}$=2.89; *P*=.06).

Mean scores for the THI in both groups and all measurements are shown in Figure 1.

Table 2. Study group results: descriptive statistics of tinnitus severity as measured with the THI subscale.

THI[a] subscale and Measurement	Range	Mean (SD)	Q1[b]	Median	Q3[c]
Functional					
Baseline	2-44	23.09 (11.67)	14.5	22	34
3 months' follow-up	0-44	18.84 (11.32)	10	17	26
6 months' follow-up	4-44	18.59 (10.15)	10	19	24
Emotional					
Baseline	4-36	19.95 (9.53)	12	20	28
3 months' follow-up	0-36	15.41 (10.26)	8	16	21.5
6 months' follow-up	0-36	14.05 (9.20)	6.5	14	18
Catastrophic					
Baseline	2-20	11.91 (4.40)	8	11	16
3 months' follow-up	0-20	9.14 (5.29)	6	10	14
6 months' follow-up	0-18	7.82 (4.88)	4	8	11.5
THI global					
Baseline	14-100	54.95 (23.98)	36.5	55	76.5
3 months' follow-up	0-96	43.39 (25.60)	20.5	41	57.5
6 months' follow-up	6-92	40.45 (22.57)	18.5	42	54

[a]THI: Tinnitus Handicap Inventory.

[b]Q1: lower quartile.

[c]Q3: upper quartile.

Table 3. Control group results: descriptive statistics of tinnitus severity as measured with the THI subscale.

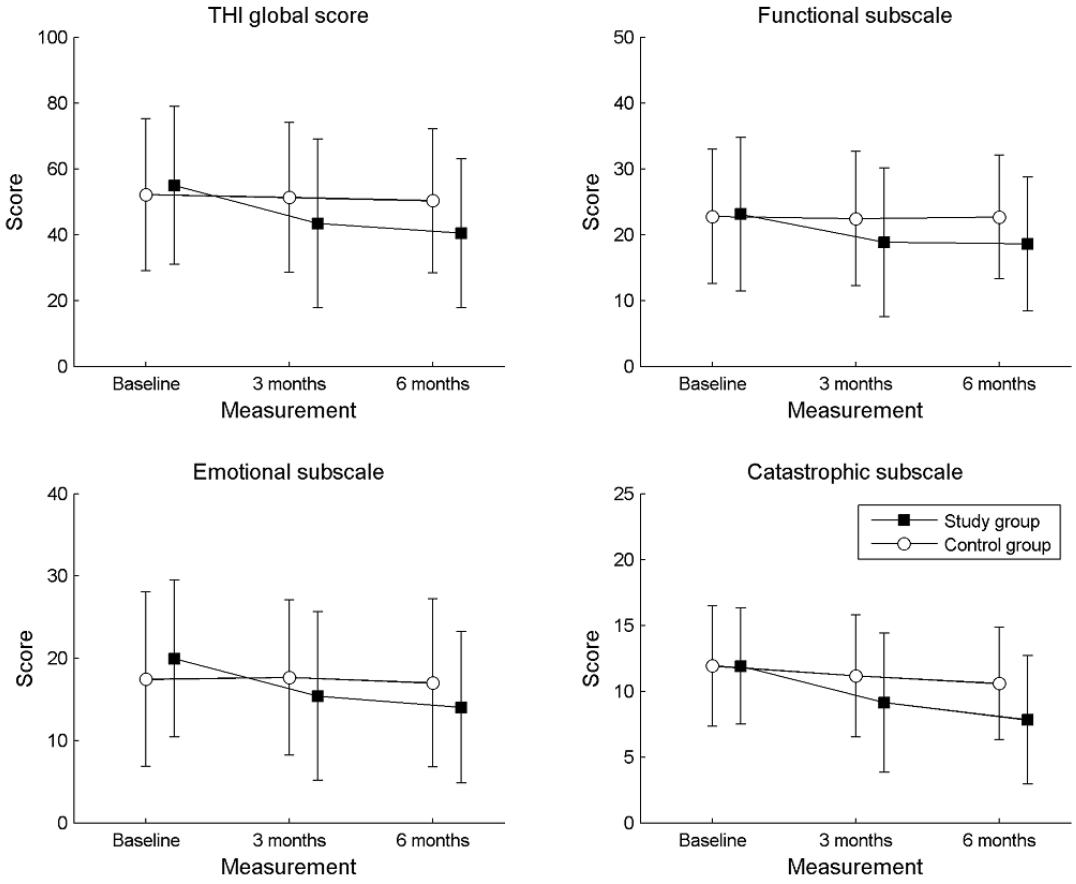
THI[a] subscale and Measurement	Range	Mean (SD)	Q1[b]	Median	Q3[c]
Functional					
Baseline	4-40	22.75 (10.18)	14.5	23	31.5
3 months' follow-up	0-40	22.42 (10.20)	18	23	28
6 months' follow-up	2-40	22.67 (9.37)	16	24	28
Emotional					
Baseline	0-34	17.46 (10.60)	8	18.5	26
3 months' follow-up	2-32	17.67 (9.41)	8.5	19	26
6 months' follow-up	0-32	17.00 (10.22)	8.5	21	25.5
Catastrophic					
Baseline	4-20	11.92 (4.59)	8	11	16
3 months' follow-up	0-18	11.17 (4.64)	6.5	12	15.5
6 months' follow-up	0-18	10.58 (4.27)	8	12	13.5
THI global					
Baseline	12-92	52.13 (23.14)	31	52.5	69.5
3 months' follow-up	6-90	51.25 (22.74)	33	54	68
6 months' follow-up	6-94	50.25 (21.82)	36.5	50	66

[a]THI: Tinnitus Handicap Inventory.

[b]Q1: lower quartile.

[c]Q3: upper quartile.

Figure 1. Mean scores obtained by the participants in the study and control groups on the THI. The squares and circles are mean scores, the error bars are SDs. THI: Tinnitus Handicap Inventory.



Clinically Important Change in Tinnitus Severity

At the 3-month follow-up a clinically significant improvement in tinnitus severity as shown with the THI (an improvement of ≥20 points) was found in 13/44 patients in the study group (30%) but in none of the controls. The difference between the groups was statistically significant (χ_1^2=8.77; *P*=.003). At the 6-month follow-up, a clinically significant improvement in the THI was found in 17/44 patients in the study group (39%) and again in none of the controls; the difference was statistically significant (χ_1^2=12.36; *P*<.001).

Tinnitus Loudness and Annoyance as Measured With VAS

Descriptive statistics for the tinnitus loudness and tinnitus annoyance scores for initial and follow-up measurements in both the study and control groups are shown in Tables 4 and 5, respectively.

For tinnitus loudness, the interaction effect (group × time) was statistically significant ($F_{2,126}$=3.19; *P*=.04; e^2=0.048) and multiple comparisons showed that tinnitus loudness at the 3-month follow-up had decreased significantly in comparison with baseline (*P*<.001). It was also significantly lower at the 6-month follow-up in comparison with baseline (*P*<.001), but there was no statistically significant difference (*P*>.99) between the 3- and 6-month follow-ups. Tinnitus loudness did not change significantly (*P*>.99) in the control group over 6 months. A similar but stronger effect was observed for tinnitus annoyance ($F_{2,126}$=11.28; *P*<.001; e^2=0.152). Mean scores for tinnitus loudness and tinnitus annoyance in both groups and all measurements are shown in Figure 2.

Table 4. Study group results: descriptive statistics of the results of tinnitus loudness and tinnitus annoyance as measured with VAS.

VAS[a] and Measurement	Range	Mean (SD)	Q1[b]	Median	Q3[c]
Loudness					
Baseline	20-100	64.55 (21.81)	50	66.5	80.8
3 months' follow-up	11-95	53.11 (23.44)	35.5	50.5	74.3
6 months' follow-up	0-95	51.66 (24.98)	32.8	51.5	75.5
Annoyance					
Baseline	10-100	64.20 (24.72)	50.5	66.5	81
3 months' follow-up	0-97	45.43 (27.13)	23.3	45	67.3
6 months' follow-up	0-95	42.61 (29.45)	17.8	39.5	66.5

[a]VAS: visual analog scale.

[b]Q1: lower quartile.

[c]Q3: upper quartile.

Table 5. Control group results: descriptive statistics of the results of tinnitus loudness and tinnitus annoyance as measured with VAS.

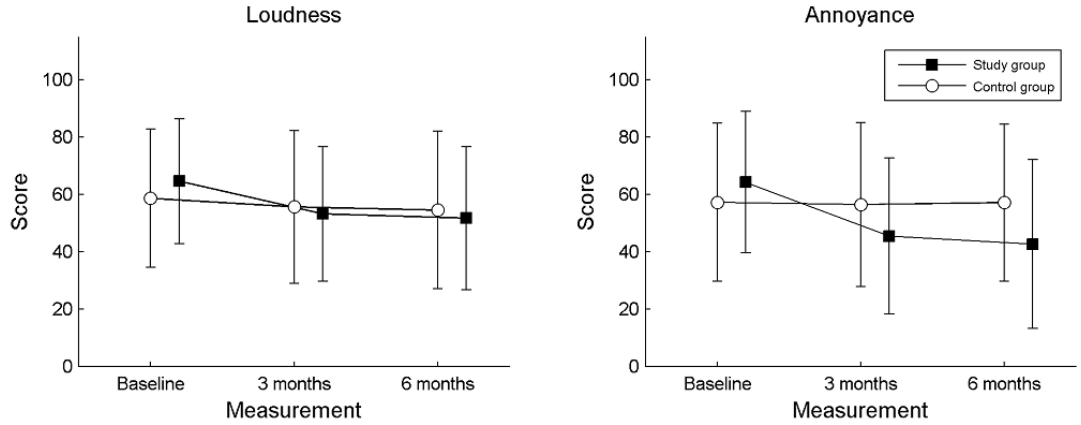
VAS[a] and Measurement	Range		Mean (SD)	Q1[b]	Median	Q3[c]
Loudness						
Baseline	14	95	58.62 (24.13)	36	60	80
3 months' follow-up	6	98	55.57 (26.70)	37	51	80
6 months' follow-up	5	92	54.50 (27.47)	31.3	58.5	78
Annoyance						
Baseline	3	97	57.19 (27.54)	38	60	80
3 months' follow-up	3	96	56.39 (28.58)	35	56	82
6 months' follow-up	3	91	57.04 (27.40)	35.5	63	80.8

[a]VAS: visual analog scale.

[b]Q1: lower quartile.

[c]Q3: upper quartile.

Figure 2. Mean scores obtained by the participants in the study and control groups on tinnitus loudness and tinnitus annoyance as measured by VAS. The squares and circles are mean scores, the error bars are SDs. VAS: visual analog scale.



Data Concerning the Use of ReSound by the Participants

A majority of study group participants (39/44, 89%) said they listened to environmental sounds, 43% (19/44) chose music, and 18% (8/44) therapeutic sounds. The most popular environmental sounds were rain and ocean waves, which were chosen by more than half of the participants. Detailed data on the sounds listened to are shown in Figure 3.

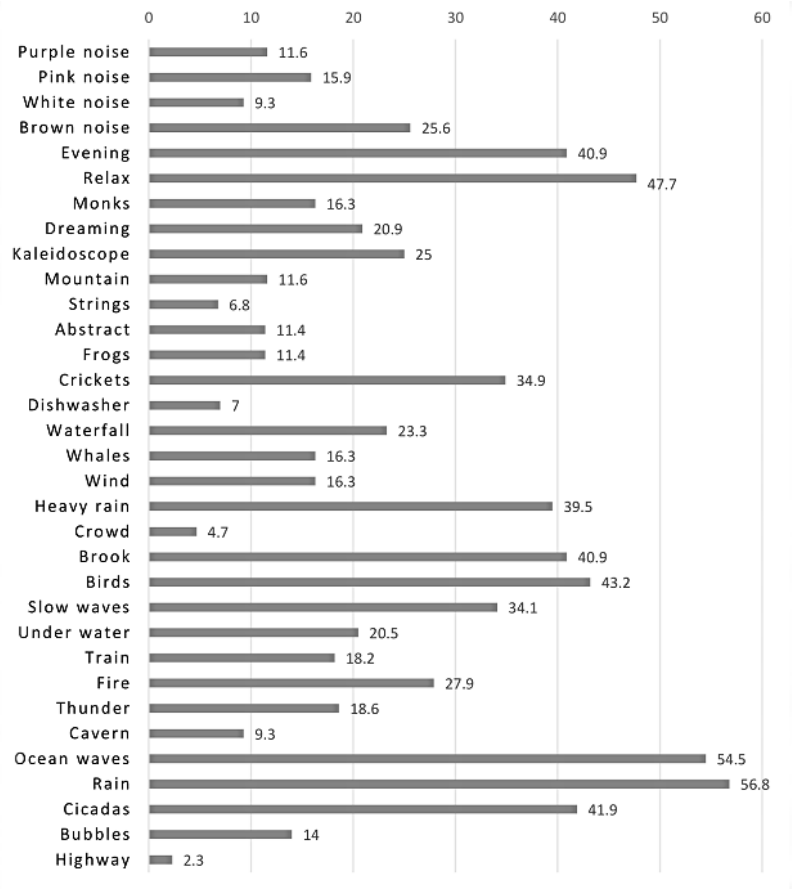
Of the 44 participants, 24 (55%) reported they used the app every day, 18 participants (41%) used it a few times a week, and 2 participants (5%) only once a week.

As many as 12 participants (27%) reported they used the app less than 30 minutes a day; 28 (64%) 1-2 hours a day, 3 (7%) 3-8 hours a day, and 1 (2%) used it for more than 8 hours a day.

As mentioned, the ReSound app also provides breathing and relaxation exercises and meditation support. Eight patients declared they had used other functions included in the app. Five of them used breathing exercises, 5 used the "Information" section, and 1 used the "Meditation" section (some people used >1 function). Therefore, in general, using additional functions was not very frequent. At the same time, participants knew that the main purpose of using the app was to listen to sounds, and they answered the surveys and questionnaires from this angle.

Participants rated the overall effectiveness of the app. Most of them (n=23, 52%) assessed it as moderate, 15 (34%) as high, 1 (2%) as very high, and 4 (9%) as low; 1 person did not answer the question.

Figure 3. Sounds listened to by the participants.



Discussion

Principal Findings

This study showed that using a mobile app that generates sounds reduces tinnitus severity. An improvement was found especially in the emotional domain. Furthermore, the patients accepted the app and its overall rating was positive.

Based on the general results of the THI questionnaire, tinnitus severity in the study group decreased significantly, whereas it remained unchanged in the control group. The overall mean tinnitus severity in the study group decreased from 54.95 to 40.45 points, whereas in the control group it remained stable (mean values: 52.13 and 50.25 points, respectively). Additionally, we found a clinically significant improvement in 17/44 patients (39%) who used the mobile app for 6 months. A similar observation was made by Henry et al [24] who tested the app in a group of 25 people with tinnitus for 8 weeks and found that almost one-third of them (8/25) reported a significant improvement. Similarly, in our pilot study [26], a clinically significant change (as calculated based on the THI questionnaire scores) was reported in 54% of our participants (n=28).

The observed reduction in tinnitus severity may be due to many mechanisms. Some were offered by Munir and Pryce [37,38], who observed that listening to pleasant sounds allowed patients to escape from a problem by providing a diversion from the intrusiveness of tinnitus. The sounds also stimulated pleasant memories and allowed a break away from reality. Another possible explanation is that the sound enrichment devices are a way of having control over tinnitus. The smartphone and its app can be regarded as a tool that enables control over an invisible condition. A study by Dauman and Dauman [39] suggested that deciding what sounds to listen to created a sense of control over tinnitus and reduced the mental effort required to distract oneself from tinnitus. They also suggested that controlling tinnitus using an external tool had a positive effect on self-empowerment and improved psychological functioning of the patient [39].

Our results show that the highest improvement in tinnitus severity was found on the Emotional subscale. This subscale identifies emotions and states that accompany tinnitus such as anger, frustration, worry, irritability, and perhaps depressive symptoms [40]. A statistically significant improvement was also found on the Catastrophic reaction subscale, which assesses reactions a patient typically encounters when tinnitus arises: feelings of desperation, enslavement, seeing it as a terrible disease, and the sense of lack of control [40]. A reduction in the emotional burden of tinnitus also may be reflected in the changes measured by the second tool, the VAS. We saw a significant improvement (VAS scores) in both loudness and annoyance in the study group, but the effect was stronger for annoyance than for loudness. By comparison, in the control group, both measures were stable over 6 months. It should be noted that in our study a change in tinnitus severity was not found for the Functional subscale. This means that the participants always felt a limitation due to their tinnitus in the areas of cognitive, social, and physical functioning. The advantage of the app was that it led to lower negative emotions and thus reduced overall tinnitus severity.

When the patients' preferences for the type of sound to listen to were assessed, almost 90% (39/44) listened to environmental sounds. The most popular sounds were rain and ocean waves. These results are consistent with those of Henry et al [24] who found that participants preferred the sounds of ocean waves and rain. In the study by Tyler et al [25], the sounds most frequently listened to and enjoyed were rain and waves on rocks, and for Perreau et al [28] heavy rain, pink noise, and waterfalls.

The usage survey showed that most of the participants used the app every day and usually for 1-2 hours a day. However, the data were based only on self-reports and we were unable to confirm them. In its present form, the app does not allow the duration of use to be recorded. This is a limitation that might be overcome in future research by providing a data logging function that could track and store user activity. Another limitation of the study is the appreciable number of participants who dropped out. The data were collected by postal mail and coincided with the outbreak of COVID-19 [41]. The introduction of a national lockdown and other restrictions reduced the ease with which patients could send in data. In the future, it would be better to collect data electronically, for example, using an online platform that ensures a full set of results.

It should be also mentioned that when treatment is applied, patients could be influenced by placebo effects. Therefore, it is worth considering the use of a technical placebo for the control group in future research. Unfortunately, in the present setting it was not possible. The app used is a commercial product, and we did not have the possibility to introduce any changes to the generated sounds.

Taken together, however, using a mobile app for tinnitus sound therapy does seem to be a promising solution. In future work, research in a more controlled environment is needed. App developers might consider implementing a system that allows data to be logged, as well as tracking the time and mode of app usage. This would enable a better understanding of how patients interact with the app and document their patterns of usage.

Conclusions

Use of an app that generates background sounds appears to be an effective way of reducing tinnitus severity. Although participants still experienced limitations caused by tinnitus in the areas of cognitive, social, and physical function, the advantage of the app was that it led to lower negative emotions and thus reduced overall tinnitus severity. It is worth considering whether mobile apps might be usefully employed to manage tinnitus in a professional setting.

Authors' Contributions

JJK conceived and planned the study. JJK, DR-K, and PHS collected the data. EG performed data analysis. JJK, EG, and WWJ wrote the manuscript in consultation with PHS. All authors approved the final version.

Conflicts of Interest

None declared.

References

1. Kim JA. Creative Destruction of Medicine: How the Digital Revolution Will Create Better Health Care. Healthc Inform Res 2013;19(3):229. [doi: 10.4258/hir.2013.19.3.229]
2. Furmankiewicz M, Sołtysik-Piorunkiewicz A, Ziuziański P. Systemy mobilne w e-zdrowiu. Studia Ekonomiczne. Zeszyty Naukowe. Yadda. 2016. URL: http://yadda.icm.edu.pl/yadda/element/bwmeta1.element.cejsh-201a710b-6d09-4085-b254-5fbe956ebc25 [accessed 2022-04-06]
3. Faragli A, Abawi D, Quinn C, Cvetkovic M, Schlabs T, Tahirovic E, et al. The role of non-invasive devices for the telemonitoring of heart failure patients. Heart Fail Rev 2021 Sep;26(5):1063-1080 [FREE Full text] [doi: 10.1007/s10741-020-09963-7] [Medline: 32338334]
4. Balsa J, Félix I, Cláudio AP, Carmo MB, Silva ICE, Guerreiro A, et al. Usability of an Intelligent Virtual Assistant for Promoting Behavior Change and Self-Care in Older People with Type 2 Diabetes. J Med Syst 2020 Jun 13;44(7):130. [doi: 10.1007/s10916-020-01583-w] [Medline: 32533367]

5. Vodička S, Naji H, Zelko E. The Role of Telecardiology in Dealing with Patients with Cardiac Rhythm Disorders in Family Medicine - Systematic Review. Zdr Varst 2020 Jun;59(2):108-116 [FREE Full text] [doi: 10.2478/sjph-2020-0014] [Medline: 32952710]
6. Długaszek M, Gumprecht J, Berdzik-Kalarus S, Chodkowski A, Nabrdalik K. Telemedicine in response to challenges of modern diabetology. Clinical Diabetology 2016;5(1):22-25 [FREE Full text] [doi: 10.5603/DK.2016.0004]
7. Fieux M, Duret S, Bawazeer N, Denoix L, Zaouche S, Tringali S. Telemedicine for ENT: Effect on quality of care during Covid-19 pandemic. Eur Ann Otorhinolaryngol Head Neck Dis 2020 Sep;137(4):257-261 [FREE Full text] [doi: 10.1016/j.anorl.2020.06.014] [Medline: 32624390]
8. Baguley D, McFerran D, Hall D. Tinnitus. The Lancet 2013 Nov 9;382(9904):1600-1607 [FREE Full text] [doi: 10.1016/s0140-6736(13)60142-7]
9. Jastreboff PJ. Phantom auditory perception (tinnitus): mechanisms of generation and perception. Neurosci Res 1990 Aug;8(4):221-254. [doi: 10.1016/0168-0102(90)90031-9] [Medline: 2175858]
10. Cima RFF, Mazurek B, Haider H, Kikidis D, Lapira A, Noreña A, et al. A multidisciplinary European guideline for tinnitus: diagnostics, assessment, and treatment. HNO 2019 Mar;67(Suppl 1):10-42. [doi: 10.1007/s00106-019-0633-7] [Medline: 30847513]
11. Møller A. Epidemiology of tinnitus in adults. In: Møller A, Langguth B, De Ridder D, Kleinjung T, editors. Textbook of Tinnitus. New York, NY: Springer; 2011:29-37.
12. Tunkel DE, Bauer CA, Sun GH, Rosenfeld RM, Chandrasekhar SS, Cunningham ER, et al. Clinical practice guideline: tinnitus. Otolaryngol Head Neck Surg 2014 Oct;151(2 Suppl):S1-S40. [doi: 10.1177/0194599814545325] [Medline: 25273878]
13. Probst T, Pryss RC, Langguth B, Spiliopoulou M, Landgrebe M, Vesala M, et al. Outpatient Tinnitus Clinic, Self-Help Web Platform, or Mobile Application to Recruit Tinnitus Study Samples? Front Aging Neurosci 2017;9:113 [FREE Full text] [doi: 10.3389/fnagi.2017.00113] [Medline: 28484389]
14. Greenwell K, Sereda M, Coulson N, El Refaie A, Hoare DJ. A systematic review of techniques and effects of self-help interventions for tinnitus: Application of taxonomies from health psychology. Int J Audiol 2016 Jul;55 Suppl 3:S79-S89. [doi: 10.3109/14992027.2015.1137363] [Medline: 27146036]
15. Malouff JM, Noble W, Schutte NS, Bhullar N. The effectiveness of bibliotherapy in alleviating tinnitus-related distress. J Psychosom Res 2010 Mar;68(3):245-251. [doi: 10.1016/j.jpsychores.2009.07.023] [Medline: 20159209]
16. Kutyba J, Jedrzejczak WW, Gos E, Bieńkowska K, Raj-Koziak D, Skarżyński PH. Self-help interventions chosen by subjects with chronic tinnitus - a retrospective study of clinical patients. Int J Audiol 2021 Aug 19:1-6. [doi: 10.1080/14992027.2021.1964040] [Medline: 34410211]
17. Barros Suzuki FAD, Suzuki FA, Yonamine FK, Onishi ET, Penido NO. Effectiveness of sound therapy in patients with tinnitus resistant to previous treatments: importance of adjustments. Braz J Otorhinolaryngol 2016;82(3):297-303 [FREE Full text] [doi: 10.1016/j.bjorl.2015.05.009] [Medline: 26541232]
18. Hoare DJ, Searchfield GD, El Refaie A, Henry JA. Sound therapy for tinnitus management: practicable options. J Am Acad Audiol 2014 Jan;25(1):62-75. [doi: 10.3766/jaaa.25.1.5] [Medline: 24622861]
19. Sereda M, Xia J, El Refaie A, Hall D, Hoare D. Sound therapy (using amplification devices and/or sound generators) for tinnitus. Cochrane Database Syst Rev 2018 Dec 27;12:CD013094 [FREE Full text] [doi: 10.1002/14651858.CD013094.pub2] [Medline: 30589445]
20. Casale M, Costantino A, Rinaldi V, Forte A, Grimaldi M, Sabatino L, et al. Mobile applications in otolaryngology for patients: An update. Laryngoscope Investig Otolaryngol 2018 Dec;3(6):434-438 [FREE Full text] [doi: 10.1002/lio2.201] [Medline: 30599026]
21. Deshpande AK, Shimunova T. A Comprehensive Evaluation of Tinnitus Apps. Am J Audiol 2019 Sep 13;28(3):605-616. [doi: 10.1044/2019_aja-18-0135]
22. Mehdi M, Stach M, Riha C, Neff P, Dode A, Pryss R, et al. Smartphone and Mobile Health Apps for Tinnitus: Systematic Identification, Analysis, and Assessment. JMIR Mhealth Uhealth 2020 Aug 18;8(8):e21767 [FREE Full text] [doi: 10.2196/21767] [Medline: 32808939]
23. Sereda M, Smith S, Newton K, Stockdale D. Mobile Apps for Management of Tinnitus: Users' Survey, Quality Assessment, and Content Analysis. JMIR Mhealth Uhealth 2019 Jan 23;7(1):e10353 [FREE Full text] [doi: 10.2196/10353] [Medline: 30672739]
24. Henry JA, Thielman E, Zaugg T, Kaelin C, Choma C, Chang B, et al. Development and field testing of a smartphone "App" for tinnitus management. Int J Audiol 2017 Oct;56(10):784-792. [doi: 10.1080/14992027.2017.1338762] [Medline: 28669224]
25. Tyler RS, Owen RL, Bridges J, Gander PE, Perreau A, Mancini PC. Tinnitus Suppression in Cochlear Implant Patients Using a Sound Therapy App. Am J Audiol 2018 Sep 12;27(3):316-323. [doi: 10.1044/2018_AJA-17-0105] [Medline: 30105356]
26. Kutyba J, Gos E, Jędrzejczak WW, Raj-Koziak D, Karpiesz L, Niedziałek I, et al. Effectiveness of tinnitus therapy using a mobile application. Eur Arch Otorhinolaryngol 2022 Mar;279(3):1257-1267 [FREE Full text] [doi: 10.1007/s00405-021-06767-9] [Medline: 33783597]

27. Raj-Koziak D, Gos E, Kutyba J, Skarzynski H, Skarzynski PH. Decreased Sound Tolerance in Tinnitus Patients. Life (Basel) 2021 Jan 26;11(2):51-56 [FREE Full text] [doi: 10.3390/life11020087] [Medline: 33530474]
28. Perreau AE, Tyler RS, Frank V, Watts A, Mancini PC. Use of a Smartphone App for Cochlear Implant Patients With Tinnitus. Am J Audiol 2021 Sep 10;30(3):676-687. [doi: 10.1044/2021_AJA-20-00195] [Medline: 34314254]
29. Newman CW, Jacobson GP, Spitzer JB. Development of the Tinnitus Handicap Inventory. Arch Otolaryngol Head Neck Surg 1996 Feb;122(2):143-148. [doi: 10.1001/archotol.1996.01890140029007] [Medline: 8630207]
30. Newman CW, Sandridge SA, Jacobson GP. Psychometric adequacy of the Tinnitus Handicap Inventory (THI) for evaluating treatment outcome. J Am Acad Audiol 1998 Apr;9(2):153-160. [Medline: 9564679]
31. Skarzynski PH, Raj-Koziak D, J Rajchel J, Pilka A, Wlodarczyk AW, Skarzynski H. Adaptation of the Tinnitus Handicap Inventory into Polish and its testing on a clinical population of tinnitus sufferers. Int J Audiol 2017 Oct;56(10):711-715. [doi: 10.1080/14992027.2017.1319080] [Medline: 28537137]
32. Skarżyński PH, Rajchel JJ, Gos E, Dziendziel B, Kutyba J, Bieńkowska K, et al. A revised grading system for the Tinnitus Handicap Inventory based on a large clinical population. Int J Audiol 2020 Jan 12;59(1):61-67. [doi: 10.1080/14992027.2019.1664778] [Medline: 31608728]
33. Wewers ME, Lowe NK. A critical review of visual analogue scales in the measurement of clinical phenomena. Res Nurs Health 1990 Aug;13(4):227-236. [doi: 10.1002/nur.4770130405] [Medline: 2197679]
34. Atan T, Atan D, Özel S. Effectiveness of Kinesio taping in the treatment of somatosensory tinnitus: A randomized controlled trial. Complement Ther Clin Pract 2020 May;39:101100. [doi: 10.1016/j.ctcp.2020.101100] [Medline: 32379629]
35. Ciminelli P, Machado S, Palmeira M, Coutinho E, Sender D, Nardi A. Dorsomedial Prefrontal Cortex Repetitive Transcranial Magnetic Stimulation for Tinnitus: Promising Results of a Blinded, Randomized, Sham-Controlled Study. Ear Hear 2020 Jul 02;42(1):12-19. [doi: 10.1097/AUD.0000000000000908] [Medline: 32639254]
36. Raj-Koziak D, Gos E, Swierniak W, Rajchel JJ, Karpiesz L, Niedzialek I, et al. Visual Analogue Scales as a Tool for Initial Assessment of Tinnitus Severity: Psychometric Evaluation in a Clinical Population. Audiol Neurootol 2018;23(4):229-237 [FREE Full text] [doi: 10.1159/000494021] [Medline: 30439712]
37. Munir S, Pryce H. How can sound generating devices support coping with tinnitus? Int J Audiol 2021 Apr;60(4):312-318. [doi: 10.1080/14992027.2020.1827307] [Medline: 33000652]
38. Pryce H, Munir S. The Role of Sound Therapies in Tinnitus Care. The Hearing Journal 2021 Apr 02;74(3):14-15 [FREE Full text] [doi: 10.1097/01.hj.0000737568.97962.1e]
39. Dauman N, Dauman R. An Empowerment Model for Individuals with Chronic Tinnitus. Ear Hear 2020 Sep 24;42(2):425-442. [doi: 10.1097/AUD.0000000000000946] [Medline: 32976248]
40. Mazurek B, Szczepek AJ, Hebert S. Stress and tinnitus. HNO 2015 Apr;63(4):258-265. [doi: 10.1007/s00106-014-2973-7] [Medline: 25862619]
41. Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, et al. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). Int J Surg 2020 Apr;76:71-76 [FREE Full text] [doi: 10.1016/j.ijsu.2020.02.034] [Medline: 32112977]

Abbreviations

TFI: Tinnitus Functional Index
THI: Tinnitus Handicap Inventory
VAS: visual analog scale

Edited by L Buis; submitted 12.09.21; peer-reviewed by S Hatzopoulos, G Searchfield; comments to author 08.10.21; revised version received 25.11.21; accepted 18.02.22; published 21.04.22.

Please cite as:
Kutyba JJ, Jędrzejczak WW, Gos E, Raj-Koziak D, Skarzynski PH
Chronic Tinnitus and the Positive Effects of Sound Treatment via a Smartphone App: Mixed-Design Study
JMIR Mhealth Uhealth 2022;10(4):e33543
URL: https://mhealth.jmir.org/2022/4/e33543
doi:10.2196/33543
PMID:35451975

Original Paper

Fully Automated Wound Tissue Segmentation Using Deep Learning on Mobile Devices: Cohort Study

Dhanesh Ramachandram[1], PhD; Jose Luis Ramirez-GarciaLuna[1*], MD, PhD; Robert D J Fraser[1*], MN, RN; Mario Aurelio Martínez-Jiménez[2], MD, PhD; Jesus E Arriaga-Caballero[2], MD; Justin Allport[1], BEng

[1]Swift Medical Inc, Toronto, ON, Canada

[2]Department of Surgery, Universidad Autonoma de San Luis Potosi, San Luis Potosi, Mexico

[*]these authors contributed equally

Corresponding Author:
Dhanesh Ramachandram, PhD
Swift Medical Inc
Suite 500
1 Richmond St W
Toronto, ON, M5H 3W4
Canada
Phone: 1 888 755 2565
Email: dhanesh@swiftmedical.io

Abstract

Background: Composition of tissue types within a wound is a useful indicator of its healing progression. Tissue composition is clinically used in wound healing tools (eg, Bates-Jensen Wound Assessment Tool) to assess risk and recommend treatment. However, wound tissue identification and the estimation of their relative composition is highly subjective. Consequently, incorrect assessments could be reported, leading to downstream impacts including inappropriate dressing selection, failure to identify wounds at risk of not healing, or failure to make appropriate referrals to specialists.

Objective: This study aimed to measure inter- and intrarater variability in manual tissue segmentation and quantification among a cohort of wound care clinicians and determine if an objective assessment of tissue types (ie, size and amount) can be achieved using deep neural networks.

Methods: A data set of 58 anonymized wound images of various types of chronic wounds from Swift Medical's Wound Database was used to conduct the inter- and intrarater agreement study. The data set was split into 3 subsets with 50% overlap between subsets to measure intrarater agreement. In this study, 4 different tissue types (epithelial, granulation, slough, and eschar) within the wound bed were independently labeled by the 5 wound clinicians at 1-week intervals using a browser-based image annotation tool. In addition, 2 deep convolutional neural network architectures were developed for wound segmentation and tissue segmentation and were used in sequence in the workflow. These models were trained using 465,187 and 17,000 image-label pairs, respectively. This is the largest and most diverse reported data set used for training deep learning models for wound and wound tissue segmentation. The resulting models offer robust performance in diverse imaging conditions, are unbiased toward skin tones, and could execute in near real time on mobile devices.

Results: A poor to moderate interrater agreement in identifying tissue types in chronic wound images was reported. A very poor Krippendorff α value of .014 for interrater variability when identifying epithelization was observed, whereas granulation was most consistently identified by the clinicians. The intrarater intraclass correlation (3,1), however, indicates that raters were relatively consistent when labeling the same image multiple times over a period. Our deep learning models achieved a mean intersection over union of 0.8644 and 0.7192 for wound and tissue segmentation, respectively. A cohort of wound clinicians, by consensus, rated 91% (53/58) of the tissue segmentation results to be between fair and good in terms of tissue identification and segmentation quality.

Conclusions: The interrater agreement study validates that clinicians exhibit considerable variability when identifying and visually estimating wound tissue proportion. The proposed deep learning technique provides objective tissue identification and measurements to assist clinicians in documenting the wound more accurately and could have a significant impact on wound care when deployed at scale.

(JMIR Mhealth Uhealth 2022;10(4):e36977) doi:10.2196/36977

KEYWORDS

wound; tissue segmentation; automated tissue identification; deep learning; mobile imaging; mobile phone

Introduction

Overview

Wounds result from the breakdown in the protective function of the skin and the loss of continuity of the epithelium. Wounds can be generally categorized into acute and chronic wounds. Normal wound healing involves four overlapping stages: hemostasis, inflammation, proliferation, and remodeling. Wound closure can be observed between several weeks to several months depending on wound size and other patient factors. Although debatable, generally wounds taking >3 months to heal are considered chronic wounds [1]. Wound progress through the 4 phases of healing is generally assessed using subjective observation of changes in size and tissue types by clinicians. Improvement to these subjective measures offers potential for better assessment of healing, improved treatment selection, and potential to predict patients at risk of developing a chronic or nonhealing wound.

Estimates indicate that 40 million patients worldwide may be affected by chronic wounds. A recent study that examined the prevalence of wounds in Canada between 2011 and 2012 found that almost 4% of inpatient acute hospitalization clients, >7% of home-care clients, almost 10% of long-term care clients, and almost 30% of hospital-based continuing care clients developed compromised wounds. Chronic wound care also imposes a hefty economic burden on the national health care system. The adverse economic impact of wound care has been well studied [2,3]. For example, it is estimated that the total direct-care cost of diabetic foot ulcers to the Canadian health care system was determined to be CAD $547 million (US $546.6 million), with an average cost per case of CAD $21,371 (US $21,364). A major concern is to ensure that health care professionals provide timely and effective wound care to affected individuals. Although better treatment protocols, drugs, and tissue regeneration methods are being constantly developed, it is imperative that research into timely treatment and wound healing monitoring is pursued in parallel. The protocols for treatments and medication may also be dependent on accurate assessment of wound healing and wound tissue identification.

Wound assessments and measurements have long been fraught with subjectivity and considerable variability between clinicians [4]. Although there has been progress made in automated wound area measurements using computer vision and machine learning, the reporting of wound tissue composition and their reactive proportions is still largely subjective. When tissue compositions can be measured objectively, the results could improve the accuracy of wound healing progress monitoring, enable data-driven pressure injury staging, and better predict wound healing times. Therefore, the objectives of our work were, first, to measure the inter- and intrarater variability in manual tissue identification and quantification among a cohort of wound care clinicians. We sought to establish the extent of variability and subjectivity in manual wound tissue measurements and how this related to specific tissue types of interests. Second, we investigated if an objective assessment of tissue types (ie, size and amount) could be achieved using a machine learning model that predicts wound tissue types. The proposed model's performance is reported in terms of numerical metrics, that is, mean intersection over union (mIOU) between model prediction and the ground truth labels. Finally, we evaluated the performance of the proposed model for wound tissue segmentation as collectively judged by a cohort of wound care clinicians observing the model predictions.

In this study, we proposed a fully automated wound and tissue segmentation technique based on deep convolutional neural networks. In *wound segmentation*, the goal is to delineate the region in the image that corresponds to the wound bed, and in *tissue segmentation*; the goal is to further breakdown regions within the identified wound bed into its constituent tissue types. The proposed deep learning models have been integrated into a mobile app and allow clinicians to obtain objective measurements, thereby eliminating the guesswork associated with wound tissue identification. This objective measurement addressed 2 challenges. First, differentiating tissue types within chronic wounds, when done manually, often varies between clinicians for a variety of reasons (eg, training and experience). Second, accurately determining the proportion or quantification (ie, measurement) of tissue types is challenging for a human. The models developed could automatically detect the location of a wound in an image, delineate the accurate boundaries of the wound, determine if any of the 4 types of tissue are present within the wound bed, and finally compute their relative proportions for reporting.

Background and Related Work

For the context of this study, we aim to identify and quantify four major tissue types present in chronic wounds using deep learning: epithelial tissue, granulation tissue, slough, and eschar which are typically reported in wound assessment tools such as the Bates-Jensen Wound Assessment Tool (BWAT) [5] and to stage pressure ulcers using the National Pressure Injury Advisory Panel pressure injury staging system (Pressure Ulcer Scale for Healing [PUSH]) [6]. These tissues are present in an open wound in various color spectra when observed through a conventional imaging sensor. Epithelial tissue is observed as being pinkish or white regions that migrate from the wound margin with minimal exudate. It eventually covers the wound bed and is the final visual sign of healing. Granulation tissue is found mainly in the red spectrum. Its presence in a chronic wound indicates that regeneration is progressing well and that the wound is being properly treated. Slough is observed as a soft, yellow glutinous covering on the wound and is a type of necrotic tissue. Made up of dead cells and fibrin, a wound may be completely or partially filled with slough. It may also be fibrous or strand-like, adhering to the wound bed. Finally, because of tissue death, the surface of the wound is covered with a layer of dead or devitalized tissue (eschar) that is

frequently black or brown. Initially soft, the dead tissue can lose moisture rapidly and become dehydrated with the surface becoming hard and dry. Colliquative necrosis are a subtype of this category and are yellow in color, similar to fibrin deposits. They are produced when the necrotic tissue softens and are, therefore, of a mushy consistency. The appearance of necrosis indicates degenerative breakdown of wound tissue.

The tissue composition and their relative quantity within the wound bed are important parameters for estimating wound healing progress. For example, the PUSH score [6] was proposed for pressure injuries and consists of three parameters: length×width, exudate amount (none, light, moderate, and heavy), and tissue type (necrotic tissue, slough, granulation tissue, epithelial tissue, and closed). Each parameter was scored, and the sum of the 3 scores yielded a total wound status score, which helped classify wound severity and identify nonhealing wounds. The relative quantities of relevant wound bed tissues are subjectively determined during assessments. As human beings, we are poor in accurately judging relative proportions, and inaccurate assessments can lead to incorrect downstream tasks like wound staging and treatment.

Wound Area Measurement

There are numerous approaches [7] for wound area measurement, which vary in their accuracy and repeatability across multiple raters [8]. The most common approach would be to use a ruler and measure the width and length of a wound. This measurement does not allow accurate area measurement as wounds are not typically rectilinear in shape. The next step up could be to lay a grid pattern over the wound and mark the number of square grid boxes which overlap with the wound and thereby estimate the area.

Computer-aided approaches for wound segmentation (defining wound area) have been proposed in the past for small, controlled data sets, and their robustness on large-scale data sets has never been proven to be effective. Techniques include active contours [9], graph cuts [10], and color histograms [11] as well as machine learning approaches such as support vector machines [12,13] and artificial neural networks [14].

With the recent advances in artificial intelligence, it has now become possible to train a deep learning model to perform automatic segmentation of chronic wounds [15-17] in an end-to-end manner. These methods forego the need for image feature engineering and can automatically *learn* a hierarchy of image features required for a specific task. Despite the increasing number of papers being published for deep learning–based wound segmentation, most approaches have only been trained and tested on limited data sets often conducted in controlled settings as shown in Table 1. In addition, most of these approaches have not been demonstrated to run on mobile devices having limited computing resources—a critical factor in enabling objective electronic wound documentation at the bedside.

Table 1. Comparison of wound image data sets used for wound segmentation model training.

Study	Database used	Type	Total images in training set	Acquisition settings
Lu et al [16]	Medetec	Public	<500	Unspecified
Yadav et al [18]	Medetec	Public	77	Unspecified
Goyal et al [17]	Lancashire DFU[a] database	Proprietary	600	Controlled, DSLR[b], and flash used
Li et al [19]	Hospital+internet search	Proprietary	950	Unspecified
Chakraborty [20]	Medetec+proprietary data	Mixed	153	Unspecified
Wang et al [21]	NYU[c] wound image database	Proprietary	500	Unspecified
Scebba et al [22]	SWISSWOU, Medtec, FUSC SIH[d]	Public	<300	Unspecified
This study	Swift Wound Data Set	Proprietary	Approximately 465,000	Uncontrolled and mobile phone camera

[a]DFU: diabetic foot ulcer.

[b]DSLR: digital single-lens reflex camera.

[c]NYU: New York University.

[d]SIH: secondary intention healing.

Wound Tissue Segmentation

Wound tissue segmentation, which is a more challenging problem, has received far less attention than wound segmentation. Tissue segmentation entails a pixel-wise classification of various tissues that are found within the wound bed region. Past approaches have attempted to solve this using image patch–based color clustering or segmentation [18,20,23]. When classifiers are used, a training data set of image patches is built from a small set of wound images. The data set of image patches are then used to train a classifier to assign each patch to a specific tissue type. During inference, a wound image is first segmented from its background, and the wound region is split into smaller image patches. Each image patch is then classified by the trained model. This process is slow and typically not robust enough to handle variations in imaging conditions (eg, lighting and image angle) as is often the case in practice. Recently, deep learning techniques such as fully convolutional neural networks have been applied to wound tissue segmentation [19]. The approach is not fully automated;

wound segmentation is performed using a dynamic color thresholding in the *YbCbCr* color space to segment the wound area, and then a fully convolutional neural network [24] is used to classify wound tissue within the segmented wound region. A limited set of images were used to train the network, and mobile implementation was not reported.

Methods

Overview

The inter- and intrarater agreement for wound tissue identification by wound care clinicians was first measured to establish the degree of variability present in visual estimation of tissue proportions and labeling tissue regions in wound images. The *Swift Wound Data Set*, which to the best of our knowledge is the largest labeled chronic wound data set for both wound segmentation and tissue segmentation ever reported in the literature for training deep neural networks for wound image segmentation and tissue segmentation, is described. Subsequently, a fully automated wound and tissue segmentation approach is presented, which is based on a deep encoder-decoder convolutional neural network and trained using data from our internal *Swift Wound Data Set*. Finally, the authors discuss the results obtained for both the interrater agreement study and the proposed deep learning technique for wound tissue segmentation in depth. A diagram depicting steps in this study is presented in Figure S1A-1 in Multimedia Appendix 1.

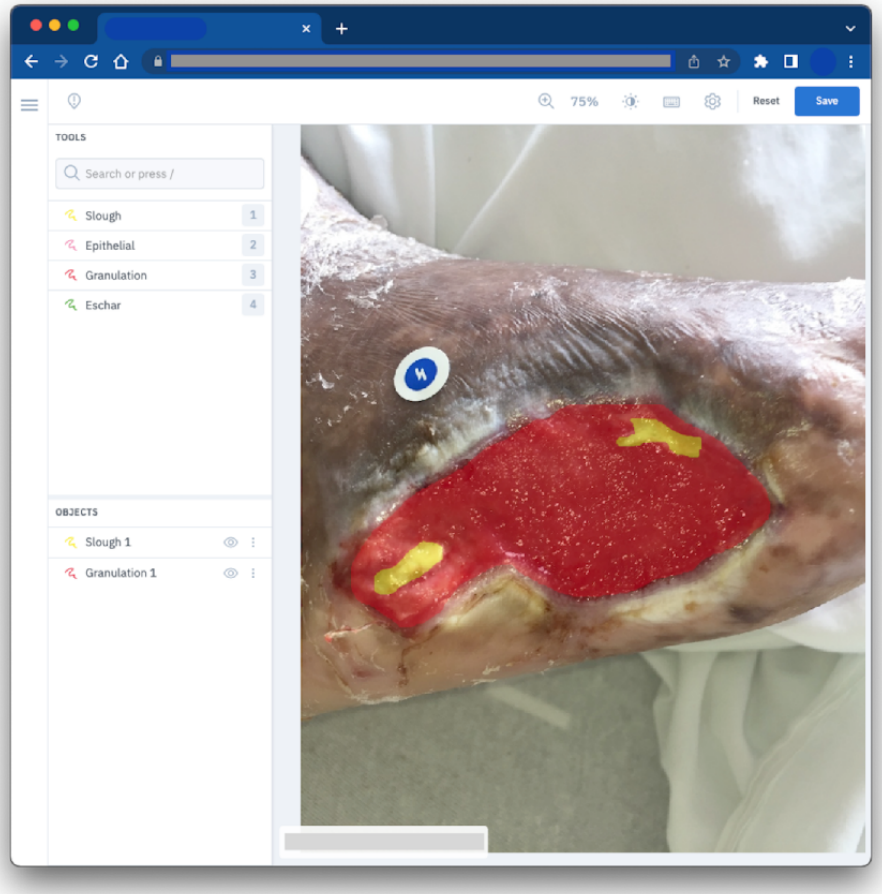
Rater Agreement in Wound Tissue Identification and Quantification

To establish the variability of wound assessment when it comes to tissue region labeling, we examined the inter- and intrarater variability between wound care clinicians when estimating not only the presence of a given tissue type in the wound bed but also their relative proportions and their confidence in the estimations. In this paper, we interchangeably use the term *raters* to refer to the wound clinicians and nurses who were involved in our study.

For this study, a random sample of 58 anonymized wound images (taken under uncontrolled lighting and viewing angles) from the *Swift Wound Data Set* consisting of pressure injuries, arterial ulcers, and venous ulcers was used. The data set was stratified according to skin tone using the Fitzpatrick scale and split into 3 subsets, with 50% overlap between subsets to measure intrarater agreement. In particular, 4 different tissue regions (epithelial, granulation, slough, and eschar) were manually labeled within the wound bed in each image. In addition, 5 experienced clinicians (a family physician, a dermatologist, a vascular surgeon, a burn surgeon, and a registered nurse) were tasked to label these images in random order using a browser-based image annotation tool shown in Figure 1. Apart from labeling (or annotating) the tissue regions, labelers were instructed to visually estimate the proportions of the 4 tissue types present within the wound bed and indicate their confidence levels when identifying these tissues.

To measure the inter- and intrarater variability, we used the Shrout and Fleiss CC [25]. As the same set of k raters labeled the same set of n samples in the data set, we used a 2-way, mixed effects model, specifically the intraclass correlation (ICC) as described in that paper to compute the ICC as a measure of reliability. Another measure of reliability is the Krippendorff α, which measures disagreement between a number of raters. For the sake of brevity, we refer readers to the study by Krippendorff [26] for a complete description of this statistical measure.

Figure 1. The web-based image annotation tool used for the interrater agreement study.



Swift Medical Wound Data Set

Previous studies on wound segmentation were either trained or validated on small wound image data sets [18-23], often acquired under very controlled conditions and focused on a limited number of wound types; for example, diabetic foot ulcers were analyzed with high-resolution digital single-lens reflex cameras with large imaging sensors (23×15.6 mm^2) and macrolenses; [17] however, questions remain as to the robustness of these approaches in real-world scenarios as results were demonstrated using very limited data.

In this study, we used our internal deidentified data set to train and validate the deep learning models for wound segmentation and tissue segmentation. Wound images in our data set were acquired using heterogeneous cellphone cameras under uncontrolled settings using the Swift Skin and Wound app from hundreds of skilled nursing facilities and long-term care centers across North America. Our data set is significantly larger (by 2 to 3 orders of magnitude) than data sets reported in previous studies [15,17,21,22] (Table 1). In addition, it is to be noted that there are no publicly available data sets for fully labeled wound tissues (ie, for epithelial, granulation, eschar, and slough), unlike the data sets listed in Table 1 which are data sets with purely binary labels (wound or background), which are much simpler to manually label.

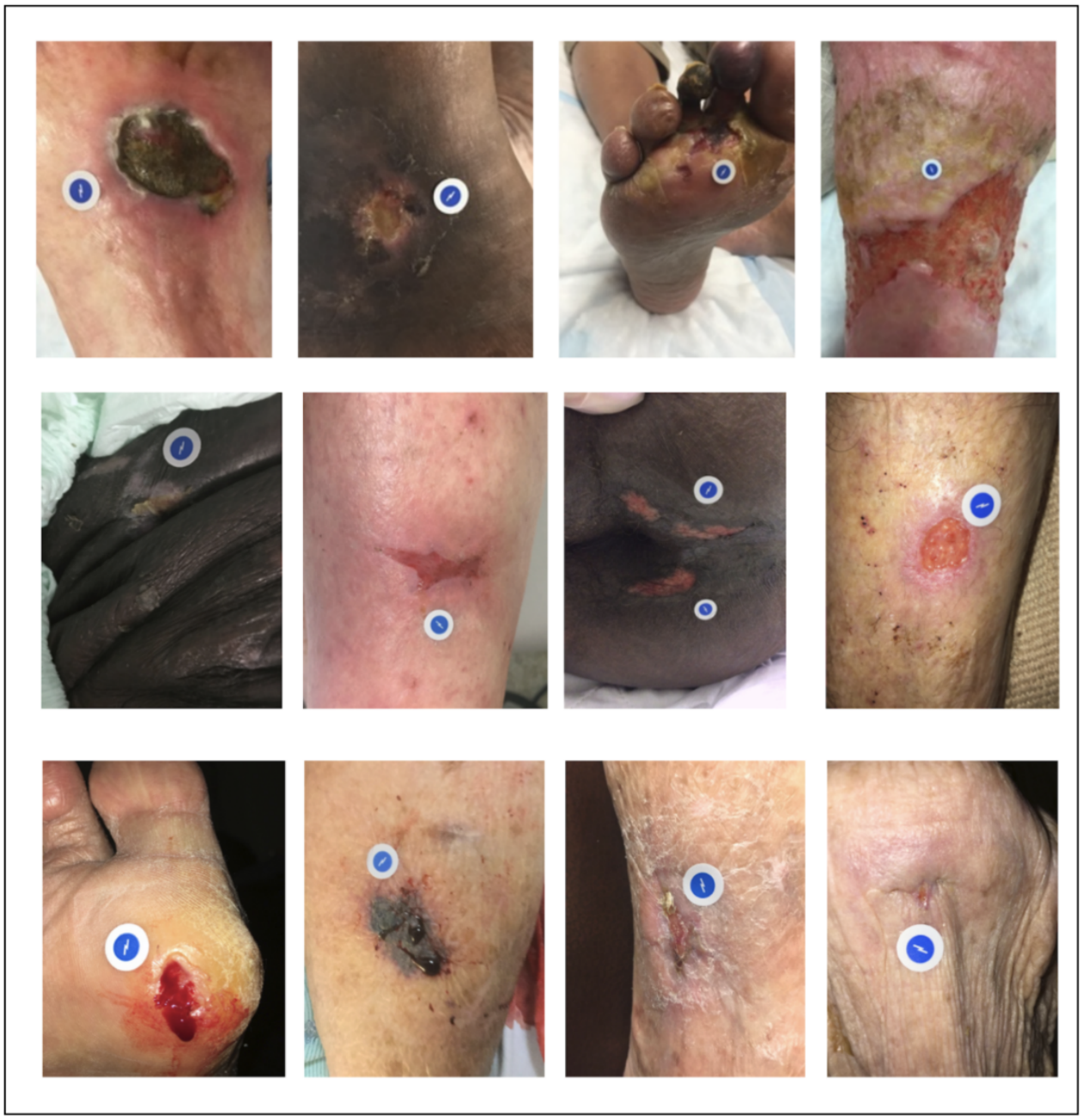
There is significant variability in terms of viewing angles, lighting conditions, background, and magnification factors for wound images in the Swift data set as shown in Figure 2. This data set also covers a wider range of skin lesions and chronic wounds than any of the previously published studies. Specifically, it consists of 14 different types of wounds or skin lesions at various stages of healing. These include bruise or abrasion, blister, burn, cancer lesion, diabetic foot ulcer, laceration, moisture associated skin damage, mole, open lesion, pressure injury, venous ulcer, rash, skin tear, and surgical wound.

There are numerous variations in skin tones because of ethnicity, which makes it challenging to isolate healthy skin and wound bed regions using traditional computer vision techniques. In most images in the data set, healthy skin area was found to be covered with age spots—a pigmentation effect associated with older individuals. Figure 2 shows examples of skin tone variations present in the data set. There is also a wide variation of the visible characteristics of the wounds in our data set; for example in terms of wound type, location, and severity.

This data set reflects the actual diversity in wound images typically observed in practice. As deep learning techniques generally scale well and perform better when trained with larger data sets, the automatic wound and tissue segmentation approach presented in this paper are expected to be more robust and accurate than previously reported approaches for large-scale

deployment. It is to be noted that owing to patient privacy concerns, we are unable to publicly share the data sets used to train our models.

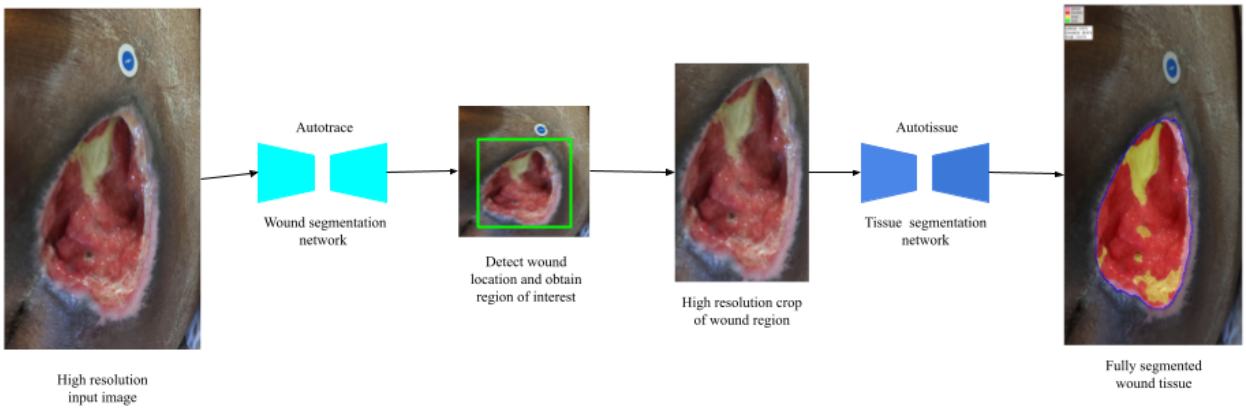
Figure 2. Sample images from the data set. The blue-white sticker seen in the images is the Food and Drug Administration–registered HealX calibrant used with the Swift Skin and Wound app for color-correction and scale calibration. Note variations in terms of viewing angles and distances, background, wound types, severity, skin tone, and wound sizes.



Fully Automated Wound Tissue Segmentation

A high-level depiction of this fully automated wound and tissue segmentation method is shown in Figure 3. A high-resolution image is first acquired using the smartphone camera. The first stage involves detecting the presence of a wound and determining the bounding box of the wound. Although other reported approaches [21] used a separate object detection model such as YoloV3 [27] to locate the wound, we used an encoder-decoder wound segmentation network, dubbed *AutoTrace*, whose predictions can not only be used to compute the bounding box of the wound, but also determine the accurate segmentation (trace) of the wound bed. This model is small and fast enough to enable real-time inference on mobile devices.

Figure 3. A high-level overview of AutoTissue, the proposed fully-automated wound and tissue segmentation approach.



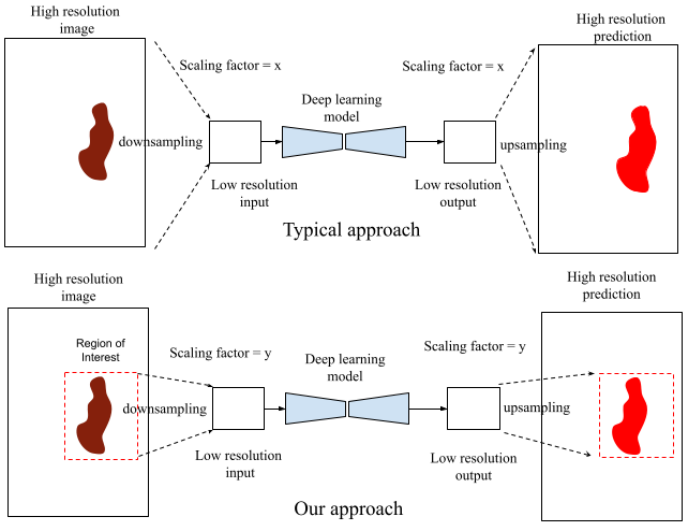
When implementing deep learning architectures, particularly on mobile devices, the input image dimension is a critical factor to ensure that memory and computation requirements are manageable. Therefore, deep learning approaches typically use a low-resolution version of the images for training and inference. A drawback of using low-resolution image inputs is that information loss is attributed to downscaling. The higher the downscaling factor, the higher the possibility of information loss owing to interpolation (see Figure 4 for an illustration). To ensure that we minimize the potential information loss from image scaling or subsampling, we take steps to apply our deep learning models on regions of interest in an image, particularly at locations where an actual wound is located within the image. Therefore, we first use the *AutoTrace* model to detect the presence of an open wound in the image, then select a bounding box encompassing the detected wound region, and finally rescale that region to the dimension required as inputs to our tissue segmentation model, that is, *AutoTissue.*

Because a wound typically only constitutes between 25% and 65% of the imaged area in our data set, applying our tissue segmentation model directly on the detected wound region ensures a high wound-to-background pixel ratio for the model inputs, which leads to more accurate predictions and less errors from the series of downsampling and upsampling operations which are applied when using the deep learning models.

The tissue segmentation network, *AutoTissue,* produces a dense prediction of 4 wound tissue types (epithelial, granulation, slough, and eschar) when present within the detected wound bed. Wound border refinement is made using the wound contour computed from *AutoTrace's* wound prediction. Here, we clip the predicted tissue predictions with the accurate wound contour to ensure only tissues that are present within the wound bed are used to compute the tissue proportions.

In the following subsections, we present a high-level overview of both the *AutoTrace* and *AutoTissue* models.

Figure 4. Diagram depicting the relationship between scaling factor and potential information loss owing to downscaling and the proposed approach. Note that scaling factor x is much larger than y as shown in the diagram.



AutoTrace: Wound Segmentation Model

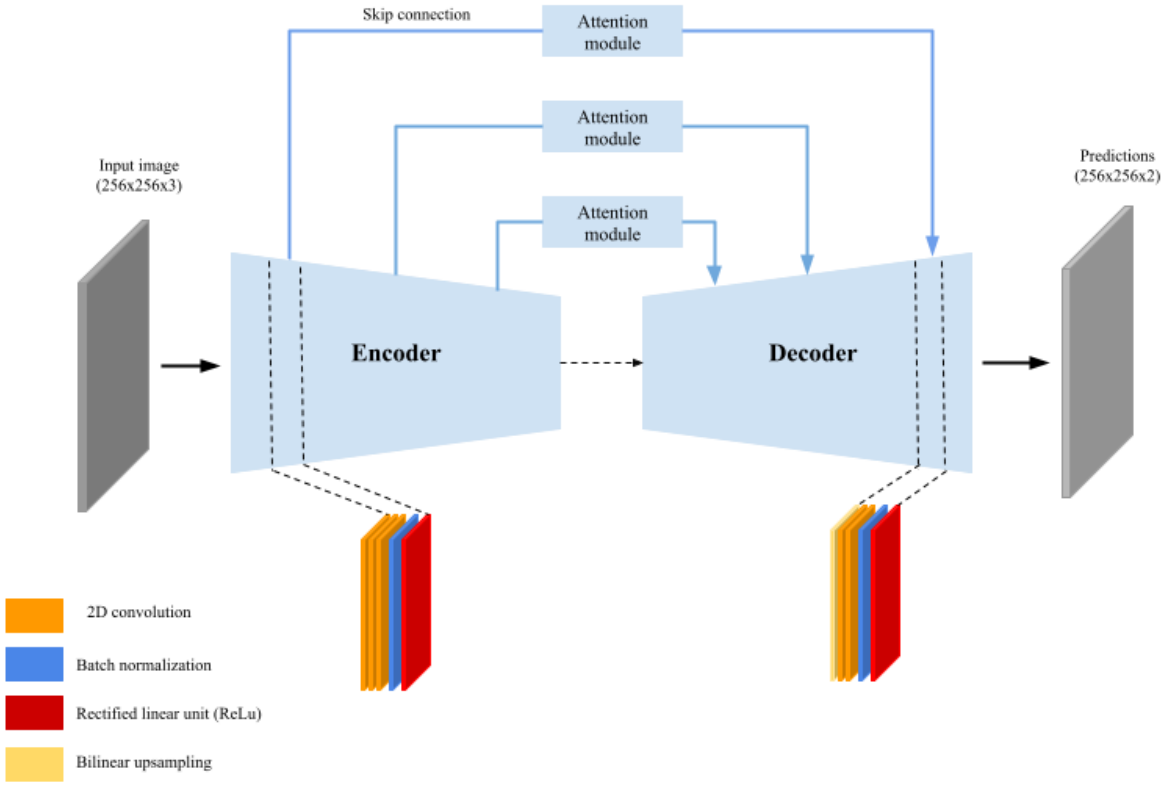
Our wound segmentation model, as depicted in Figure 5, is a deep convolutional encoder-decoder neural network with attention gates in the skip connections. The encoder block is responsible for feature extraction, and the decoder block *decodes* the learned features to produce the required output (ie, the segmentation mask). The *AutoTrace* architecture was derived from the study by Schlemper et al [28] who first proposed attention gates in convolutional neural networks. Additional customizations were implemented in our models to allow them to run on mobile devices. We replaced the normal convolutional

blocks with depth-wise separable convolutional layers [29]. The main advantage of replacing normal convolutions with depth-wise separable convolutions is the significant reduction in computation required with only a small penalty to the final accuracy. Second, we implemented strided depth-wise convolutions that can *learn* to downsample activations instead of a fixed max-pooling operation for downsampling. Third, an additive attention gate was placed in each of the skip connections in the architecture. The inclusion of additive self-attention modules in the skip connections regulates the flow of activations from earlier layers. The attention coefficients identify salient image regions and prune feature responses to preserve only the activations relevant to the specific task. This ultimately provides improved performance for the wound segmentation task. Finally, to further reduce computational and memory requirements, the decoder blocks consisted of a bilinear upsampling followed by 2 depth-wise separable convolution layers per block instead of transposed convolution layers.

We trained this model on 467,000 image-label pairs with wound region labels provided by clinicians. Our held-out test set consists of 2000 image-label pairs of arterial, venous, pressure, and diabetic ulcers taken in diverse imaging conditions and wound locations. During training, data augmentation performed included random crops, horizontal or vertical flips and random contrast and brightness adjustments. Unlike the U-Net with Attention model [28] which was trained using deep supervision, we trained our model using a single loss function by minimizing the soft dice loss which is the form of the following:

where  is the predicted probability of the pixel and  is the ground truth of the pixel. The early stopping criterion was used to stop the training after convergence. L2 regularization and dropout regularization [30] were used to control overfitting.

Figure 5. A graphical representation of the AutoTrace model for wound segmentation. ReLu: Rectified Linear Unit.



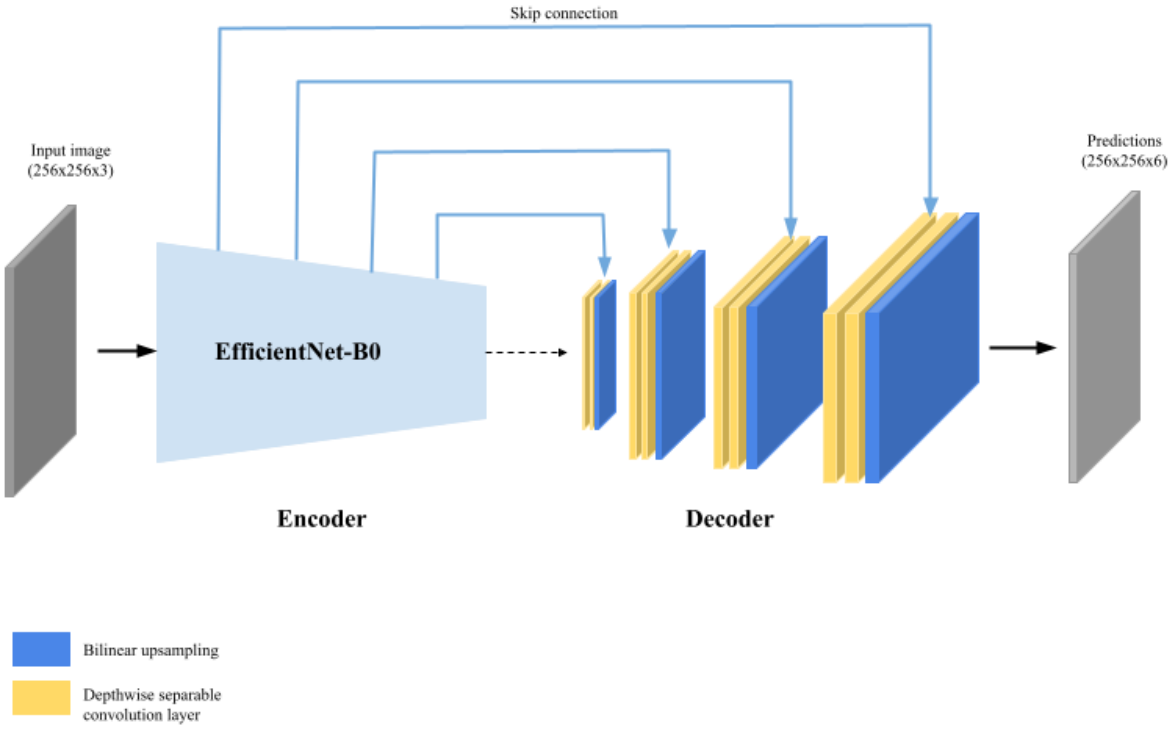
AutoTissue: Tissue Segmentation Model

The wound tissue segmentation model presented in this paper is significant as most of the previously published studies using deep learning in the domain focused on wound segmentation [15-18,20] and not tissue segmentation. The *AutoTissue* model (shown in Figure 6) is an encoder-decoder convolutional neural network that uses an *EfficientNetB0* architecture [31] as the encoder. The decoder is made up of 4 blocks; each of which consists of a single 2-dimensional bilinear upsampling layer followed by 2 depth-wise convolution layers.

The *AutoTissue* model was trained using a subset of 17,000 anonymized wound images from the *Swift Wound Data Set*

where healthy tissue, background, the HealX calibrant sticker, and the 4 wound bed tissue regions, if present, were labeled in the images. The data set was meticulously labeled by a team of trained labelers and was curated by a panel of wound clinicians. The authors could not identify any published work that used labeled data at this scale for deep learning–based wound tissue segmentation in the literature. Data augmentation, a technique used to increase the amount of training data and prevent overfitting when training deep learning models, was performed on the fly, during model training by applying random crop and rotation, random color jittering and cutout regularization [32]. Both networks were trained using AdamW (Adam With Decoupled Weight Decay) [33] adaptive learning rate using an initial learning rate of 0.001. The held-out test set consisted of 383 images consisting of stage-2 pressure, arterial, and venous ulcers and diabetic wounds.

Figure 6. Graphical representation of the AutoTissue architecture for wound tissue segmentation.



Results

Interrater Agreement Study Results

First, the authors presented the results obtained from the interrater agreement study. As mentioned earlier, the data set was split into 3 subsets, and each subset was presented to the raters (clinicians) for labeling at 1-week intervals. In all, 50% (29/58) of the images were labeled thrice, each presented 1 week apart to measure intrarater agreement. From each manually labeled image, the tissue proportions within the wound bed were assessed by counting the pixels that belonged to a certain class and computing its proportion against the total wound area. In addition, visual estimation of the tissue proportions was also recorded.

Figure 7 illustrates an example set of labels made by the wound clinicians in our study. Note the considerable variability between the labels and this observation in this example. A similar observation extends to the entire set of 58 images used to study the interrater variability and is captured by the interrater agreement ICC score presented in Table 2 Additional examples of inter- and intrarater variability in labeling tissue are provided in Multimedia Appendix 1.

The intrarater agreement for computed tissue proportions is presented in Table 3. As mentioned earlier in this section, a subset of wound images were repeatedly presented to the clinicians to label at 1-week intervals. Thus, each image has a set of 3 labels provided by the same clinician.

Figure 7. The variability between raters in labeling different tissue regions is visualized in this figure. The colors of the labels correspond to different tissue types: red corresponding to granulation, pink to epithelial, yellow to slough, and green to eschar.

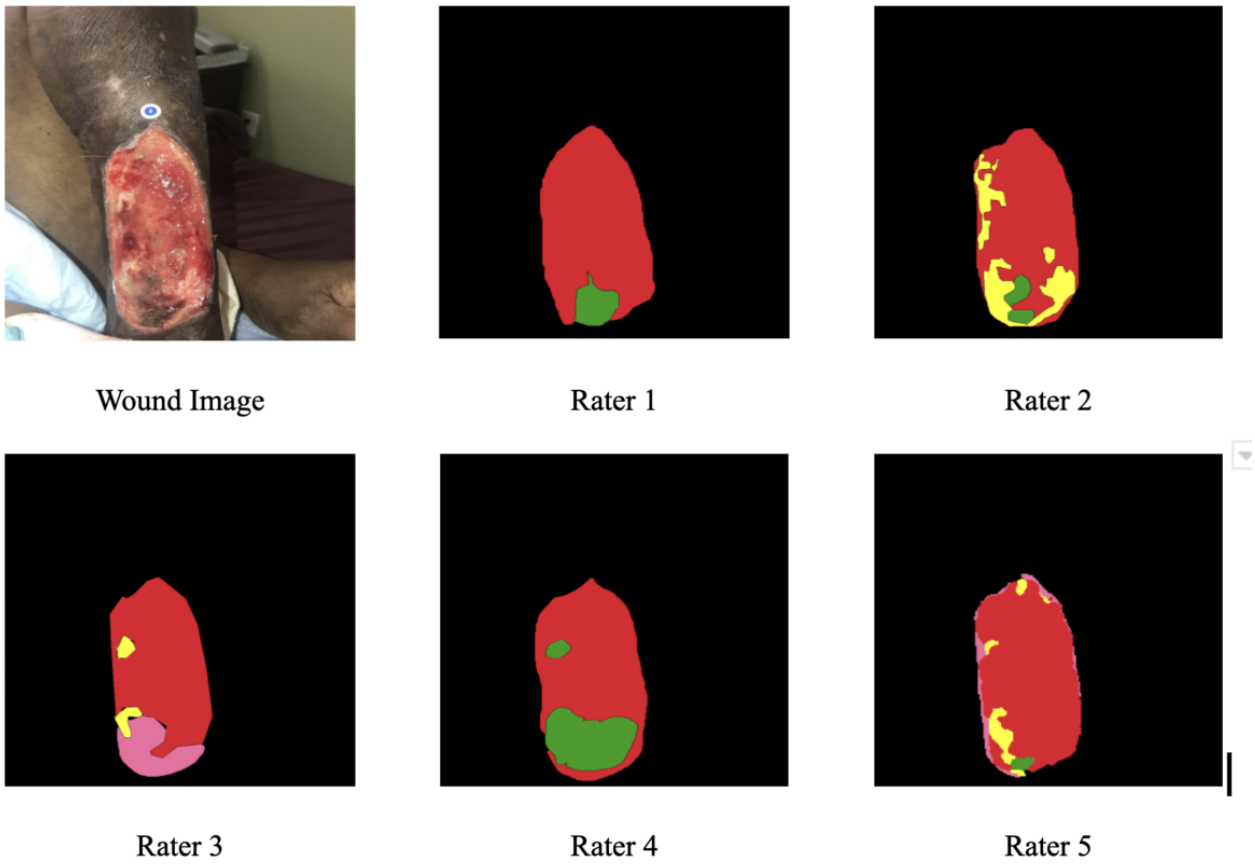


Table 2. Interrater agreement intraclass correlation for tissue proportions that were computed from wound images labeled by wound clinicians in our study.

Tissue type	Intraclass correlation
Epithelial	0.389
Granulation	0.765
Slough	0.591
Eschar	0.759

Table 3. Intrarater intraclass correlation for tissue proportions that were computed from wound images labeled by wound clinicians in our study.

Rater	Epithelial	Granulation	Slough	Eschar
Rater 1	0.785	0.789	0.843	0.803
Rater 2	0.410	0.685	0.836	0.840
Rater 3	0.535	0.729	0.641	0.493
Rater 4	0.475	0.806	0.745	0.809
Rater 5	0.757	0.958	0.986	0.963

As can be seen in Table 3, the high ICC score for individual raters (rows in the table) signifies that raters were relatively consistent when labeling (thereby reflected by computed tissue proportions) the same image multiple times over a period. The only exception was the relatively poorer intrarater agreement for epithelial tissue labeling compared with other tissue types.

Although there was moderate to high agreement in the *intrarater* agreement, only moderate *interrater* agreement was observed between raters when labeling tissue types in wound images. In particular, interrater agreement was poor for epithelial and slough tissues and moderate for other tissues based on ICC, values shown in Table 2. Note that the computation of tissue proportions as identified by experts as performed in this study

differs from how tissue proportions are *visually estimated* in practice, which can be extremely subjective. The images were labeled using a browser-based image annotation tool that allows precise annotation of different tissue types; however, in practice, wound clinicians do not have the time or tools to perform the same. We can, therefore, anticipate even higher inter- and intrarater variability in subjective visual estimations of tissue proportions compared with that reported in this study.

Figure 8 shows a box plot depicting the differences between the computed proportions based on labeled regions and the visual estimates of 4 different tissues present in the set of 58 labeled wound images as labeled in the inter- and intrarater agreement study. The subjectivity in visual estimation naturally leads to variability between the rater's visual estimates and the proportions computed from tissue labels provided by the raters through the image annotation tool. Raters largely overestimated epithelization and eschar (shown by mean and median of the box plot being negative values) and underestimated granulation and slough during visual estimation. There is substantial variability (in terms of SDs) in the distribution of errors between estimation and computed proportions for all tissue types in the range of 38% to 39%.

Apart from measuring inter- and intrarater agreement for tissue proportions calculated from labeled tissue regions, we additionally computed the interrater agreement in the clinician's ability to identify the presence of any 1 of the 4 tissue types in the wound images presented to them. As this involves binary decisions (ie, *True* when a given tissue is labeled as present and *False* when it is not—disregarding the proportions computed), we used the Krippendorff α to measure the interrater agreement. Our results, as presented in Table 4, indicate very poor interrater agreement in determining the presence and regions of epithelial tissue with a Krippendorff α value of only .014, whereas fair to moderate agreement was scored for the other tissue types. Granulation was the most agreed upon tissue type, which was in line with observations in clinical practice.

Figure 8. Box plot showing difference (in percentages) between computed tissue proportions and visual estimates for different tissue types in the rater agreement study. Negative differences indicate overestimation of rater's visual estimation. Scatter plot shows actual distribution of data points for computed differences in tissue proportions. Red triangle point indicates the mean.

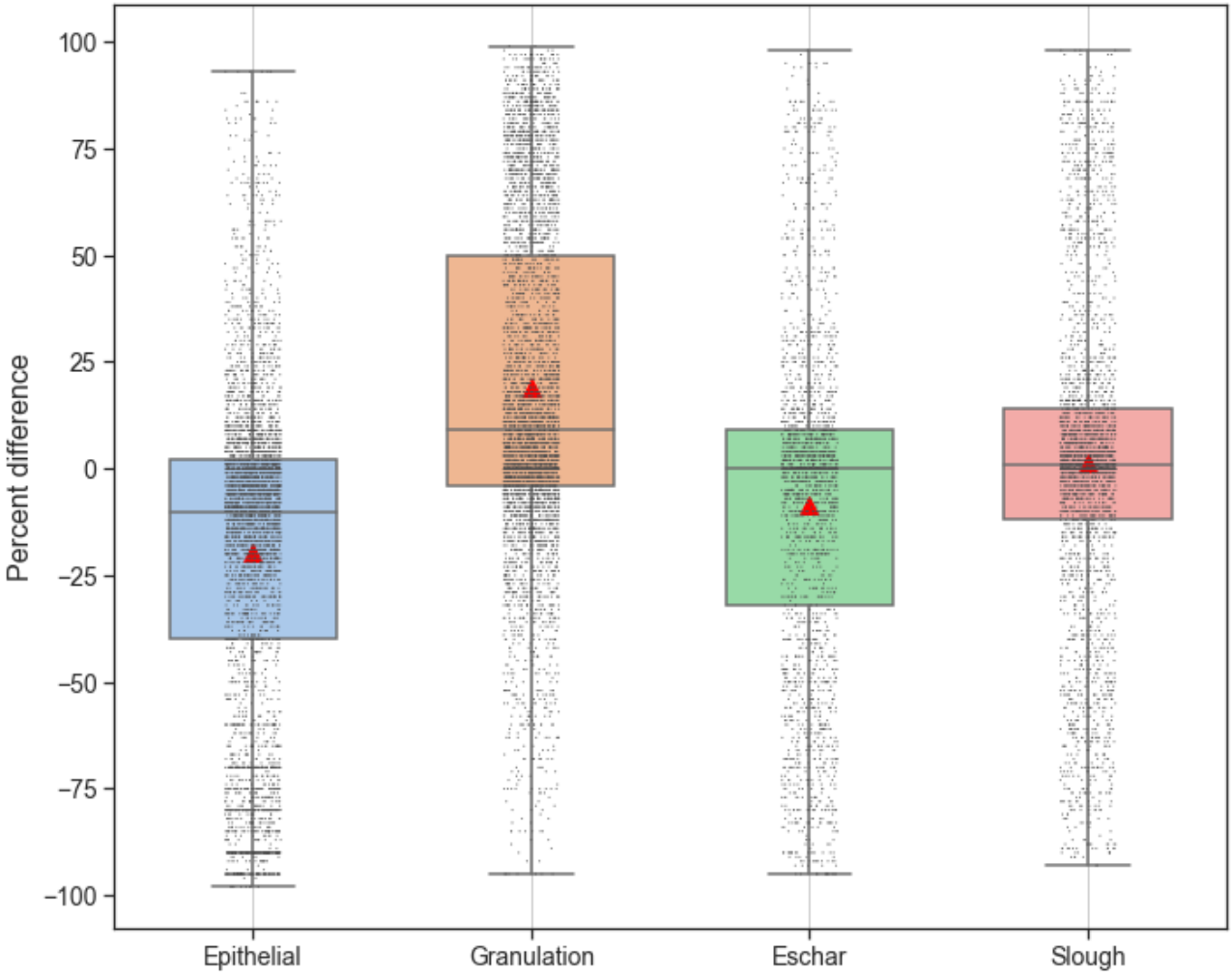
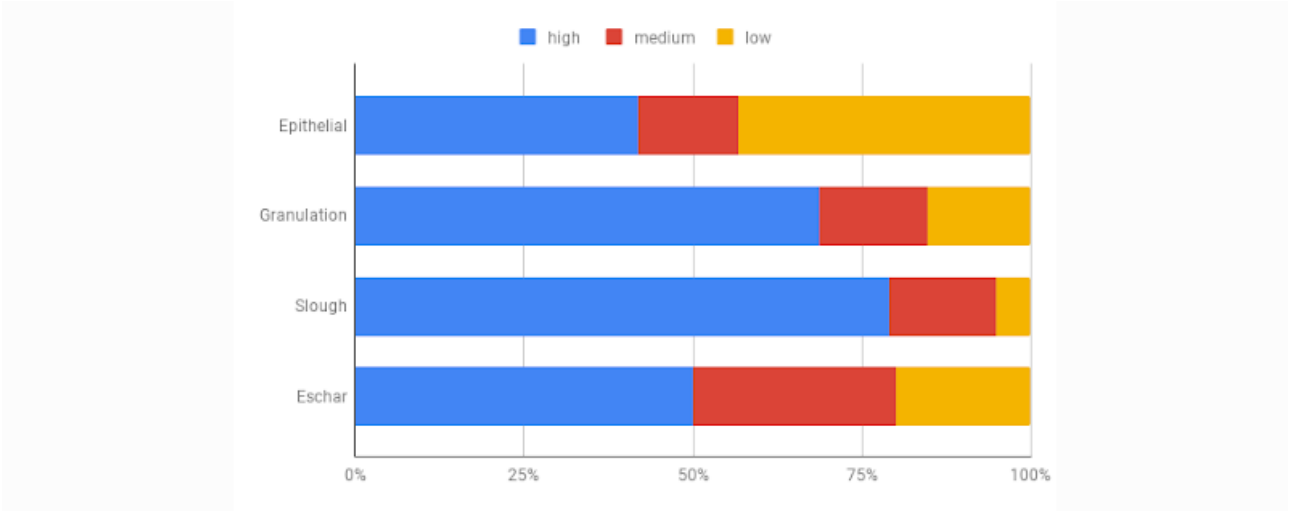


Table 4. Interrater agreement in identifying wound tissue types.

Measure of reliability	Epithelial	Granulation	Slough	Eschar
Krippendorff α	.014	.664	.415	.379

One final parameter that we captured during this study was each clinician's confidence in labeling the 4 tissue types in question. Results indicate that the clinicians involved in the study were generally very confident in labeling granulation and slough tissues, moderately confident when labeling eschar, and least confident when labeling epithelial tissue as shown in Figure 9. This result correlated well with the Krippendorff α value we observed in Table 4. The significance of this observation will be seen later when we discuss our model performance for different tissue types. Data and code pertaining to these experiments is available for public download on the web [34].

Figure 9. Clinician's confidence in tissue identification.



Automated Wound and Tissue Segmentation Results

The performances of both the wound segmentation and wound tissue segmentation models were evaluated separately on 2 different held-out test sets. We evaluated the performances of our models by computing an objective numerical metric, which is the mIOU between the ground truth labels and the predictions made by our models. The intersection over union metric measures the number of pixels common between the target and prediction label masks divided by the total number of pixels present across both label masks (see Figure 10 for a graphical depiction). When there are several classes of labels involved (eg, in the case of wound segmentation, there are two classes to be predicted, ie, wound and background classes), then the mean value of individual per class intersection over union is computed to arrive at a single metric, which is the mIOU.

The wound segmentation model, *AutoTrace*, achieves a mIOU of *0.8644* for wound region segmentation, whereas the *AutoTissue* model achieves a mIOU of *0.7192* for tissue segmentation on the held-out test sets. Several sample predictions made using our technique are presented in Figure 11. See Figure S1C in Multimedia Appendix 1 for additional results

Figure 10. A graphical representation of mean intersection over union (IOU) which varies from 0.0 (no overlap) to 1.0 (perfect overlap).

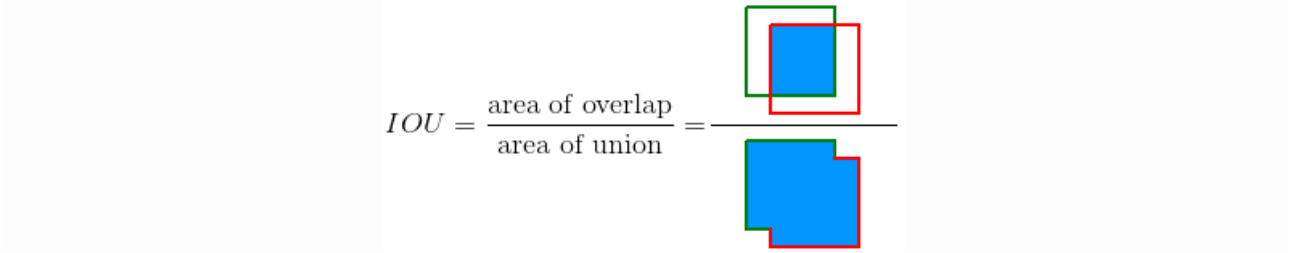
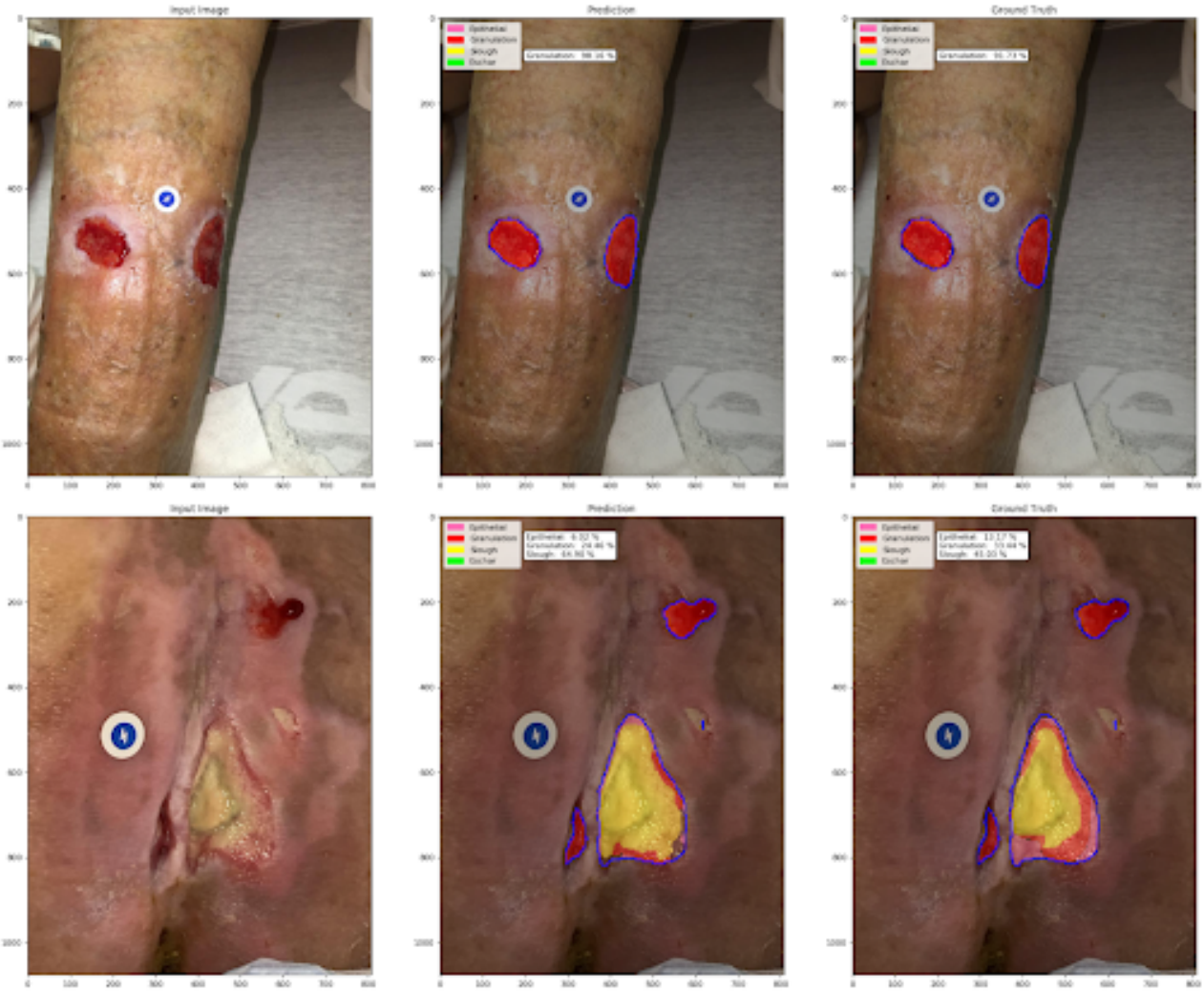


Figure 11. Sample wound and tissue segmentation results. (left to right: input image, model prediction, and ground truth). The blue contour is the wound region as determined by the AutoTrace model.



The normalized confusion matrix for the model predictions is shown in Figure 12. The confusion matrix indicates that the *AutoTissue* model is able to accurately distinguish between the wound region and healthy skin or background. Similarly, the model is performant when segmenting the HealX calibrant sticker as its appearance is relatively consistent on all images. Granulation, slough, and eschar tissue prediction performance is also favorable. We can observe that slough is largely misclassified as granulation and vice versa. This primarily reflects the challenges faced by wound clinicians when labeling regions in the wound bed that show the mixed presence of slough and granulation tissue where labelers tend to be less confident or inconsistent across different images as shown in Figure 13.

Figure 12. Normalized confusion matrix for the AutoTissue model on the held-out test set. Background (BG) includes all nonwound bed pixels including healthy tissue and background, HLX represents the calibrant sticker used for computing accurate wound measurement. EPI: epithelial; ESC: eschar; GRA: granulation; SLO: slough.

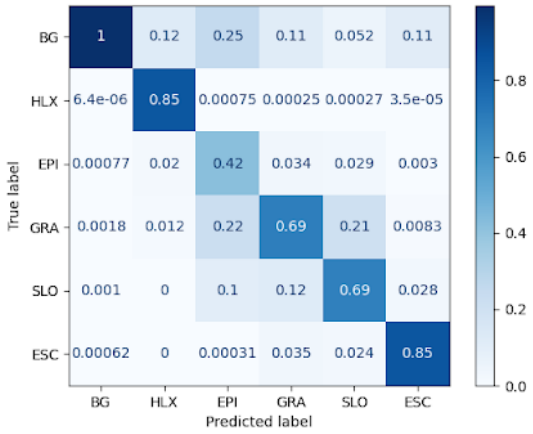


Figure 13. A wound region where pinkish tissue outside the wound bed is labeled as background in our training data set, as we are only interested in tissues within the wound bed. Note that these pinkish regions share similar appearance as tissues belonging to the epithelial class. BG: background.

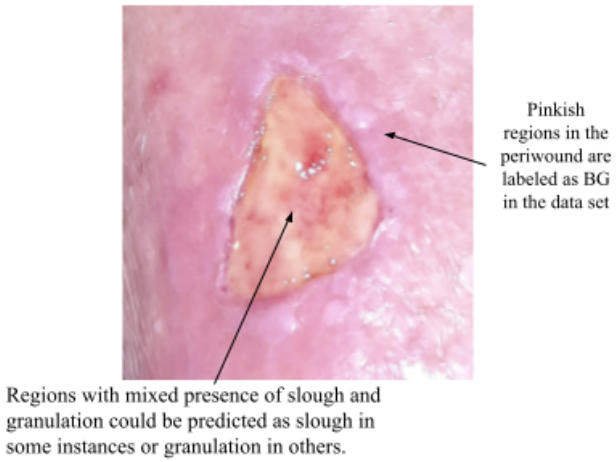


Table 5 presents the classifier report for the *AutoTissue* model. As noted, the epithelial classification exhibits low precision and low recall and a corresponding F_1-score of only 0.253. In contrast, the F_1-score for slough and eschar detection is relatively high, registering values of 0.731 and 0.802, respectively.

On the test set, the model correctly predicts 42% of pixels belonging to epithelial tissue; however, at the same time confuses epithelial tissue with healthy skin, which is part of the background or nonwound class. We attribute the model's poor performance for this class of wound tissue to the fact that the epithelial class is underrepresented in the training data set and it is challenging to label correctly within the wound bed region. Because tissue proportions are computed for regions within an open wound, epithelial tissue found in the periwound region is not considered when computing the proportions. In addition, we observed in our study that there is very poor agreement between raters in labeling epithelialization within the wound bed.

We can note that slough is sometimes misclassified as granulation and vice versa. This primarily reflects the challenges faced by wound clinicians when labeling regions in the wound bed that show the mixed presence of slough and granulation tissue where there could be considerable disagreement between raters.

Apart from testing model performance on the held-out test data as is typically reported in machine learning literature, we additionally used the 58 images which were part of the interrater agreement study to measure the agreement between predictions of the AutoTissue model and wound clinicians' labels in terms of the mIOU metric. From Table 6, we note that there is a relatively high degree of agreement for the intersection over union metric for all tissue types between our model's output and the clinicians' ground truth segmentation. In other words, there is high consistency between our model's segmentation results compared with experts' labels.

Finally, we measured a consensus-based evaluation of the correctness of the predictions made by our deep learning models. The set of 58 images and corresponding *AutoTissue* wound tissue predictions were shown to the group of wound clinicians.

We then requested the wound clinicians to collectively examine the model's predictions, discuss, and provide a quality rating based on a consensus-based agreement. This approach demonstrated that 91% (53/58) of the images were jointly rated as being very good to fair, and only 9% (5/58) of the predictions were rated as being poor. This provides an additional validation of the plausibility of our model predictions on the 4 tissue types present within the wound bed.

Table 5. Classification report for the AutoTissue model.

Metric	Epithelial	Granulation	Slough	Eschar	Average
Precision	0.180	0.623	0.783	0.759	0.586
Recall	0.424	0.693	0.685	0.850	0.663
F_1	0.253	0.656	0.731	0.802	0.610
Sensitivity	0.424	0.693	0.685	0.850	0.663
Specificity	0.603	0.772	0.862	0.825	0.765

Table 6. Interrater agreement intraclass correlation for per-tissue intersection over union between AutoTissue and expert labelers.

Tissue type	Intraclass correlation
Epithelial	0.764
Granulation	0.861
Slough	0.736
Eschar	0.855

Discussion

Principal Findings

Although we have established that there is a considerable variability in tissue labeling even between trained wound clinicians, it might appear contradictory that a deep learning model that has been trained using *noisy* labels can perform as well as humans. Arpit et al [35] suggested that sufficiently large deep neural networks did not memorize the data when trained on data sets that had mostly correct labels. Multiple studies [36-38] have also shown that a machine learning model trained using a large-scale data set of nonexpert labels can still match the performance of experts in medical image segmentation. During training, these models learn the dominant patterns observed in the data set that are shared across the data set. Therefore, to put this into the context of our own model, as the labels in our data set are pixel-wise labels (ie, there exists a label for each pixel in the image), there is an overwhelming majority of pixels that do have correct labels associated to them and a small percentage of pixels that have wrong labels associated to them owing to interrater variability. However, the *noise* that may be present in our labels is generally distributed across images and raters. Despite this, our models have the capacity to learn the dominant features for each of the tissue types and therefore are able to generalize well to unseen instances owing to the distributed and hierarchical representation, which is inherent in the design of deep neural network architectures and aided by the regularization schemes (eg, cutout and L2 regularizations) that we implement when training the models. Training on a very large data set, as was the case in our study, helped mitigate the effects of *noisy* or inaccurate pixel-wise labels. In future studies, we would want to pursue several methods [39] for dealing with noisy labels in training data to further improve our segmentation results. The labeling confidence we observed for different tissue types has direct implications to the performance of the models. We note that the model performs well for tissue classes that are easier to label and for which there are less ambiguities among labelers, and vice versa.

Still a major issue within the dermatology and medical community in general is that physicians are not trained to assess dark skin well, including wounds because most medical textbooks have illustrations that feature predominantly light-colored skin, and physicians still face a huge challenge in detecting certain tissues in dark skinned individuals. An unfortunate outcome of this implicit bias is that an accurate and prompt diagnosis may not always be possible with individuals of the Black, Indigenous, and people of color communities. For example, necrosis at the wound edge, which is an important finding, is still challenging to identify on dark skin. Our wound and tissue segmentation models on the other hand have been trained on a diverse set of wound types and skin tones, and this is a major step in ensuring the models do not inadvertently learn a bias toward a particular skin tone, which naturally leads to enhancing the ability to care for all patients.

Timely, accurate wound assessment and reporting is important for modern wound care practice. Rather than paper-based measurements and wound assessments, electronic wound assessments could have a large impact on the wound healing progress as such systems provide a more objective wound measurement, allow tracking of wound healing progress, and minimize errors or incomplete assessments. We believe that with the current technological advances, smartphone-based wound assessments will continue to have an increasing footprint in modern wound care practice. Therefore, we prioritized designing our deep learning models to be able to execute on a wide range of off-the-shelf smartphones, without the need for off-line processing. Both the *AutoTrace* and *AutoTissue* models

have been integrated into the Swift Skin and Wound app and are very performant for real-time inference on mobile devices (Multimedia Appendix 2). Both models have a combined size of <16 million parameters and a peak memory consumption of approximately 85 MB per model during inference. The inference time averaged 300 ms on mobile central processing units when tested on low-end devices and is considerably faster on newer devices that support graphics processing unit acceleration.

Limitations

The relatively poor performance of the model for epithelial tissue can be attributed to several factors, including the challenge in labeling epithelialization within the wound bed and the distinction between *epithelialization* and *epithelial tissue*, that is, intact skin resulting from healing process in the periwound region, which also appears pinkish but is not considered during labeling. While acknowledging this limitation, we believe that there would be no significant impact on pressure ulcer staging (PUSH) or wound assessment (BWAT) because epithelialization is not critical for these measures.

The Swift Skin and Wound app used for measuring and documenting wounds and the HealX fiducial markers are each Food and Drug Administration–registered Class I medical devices which are being used in over 4000 organizations across North America. We agree that verification, validation, and continued monitoring of artificial intelligence performance, including understanding the way outputs are used are critical and are the core to deployment of such models. This study documents some, but not all, of the extensive validation undertaken as part of our design, development, risk management, and regulatory processes. The deep learning models reported in this manuscript have been integrated into the Skin and Wound device app but are yet to be deployed on a wide scale in these organizations. The predictions from this model serve to assist clinicians to document wounds in an objective manner given that our inter- and intrarater studies have shown that there is a high degree of variability when humans manually detect and estimate wound tissue proportions. In the mobile app, the models' outputs serve an informative role and do not constitute diagnosis or therapy. All Swift Skin and Wound users are provided both live training and training materials, wherein these issues are addressed. We continue to monitor the performance of all models after deployment and regularly assess all our products to determine whether changes should result in reclassification or premarket notification or authorization. The models discussed in this work were assessed through our regulatory process to not cause such a change in classification.

Conclusions

Significant interrater variability in visual estimation of tissue proportions by a cohort of wound care clinicians reflects the subjective challenge of tissue typing. Epithelialization is the most varied measurement and can be linked to the challenges observed in clinical practice in identifying that tissue type. To reduce ambiguity and provide objectivity in tissue identification, we present a framework for deep learning–based wound segmentation and tissue segmentation that is capable of running in near real time on off-the-shelf smartphones. To the best of our knowledge, our models have been trained using a chronic wound image data set that is not only magnitudes larger than any previously reported data sets but also the most diverse in terms of wound types and skin tones allowing for unbiased, robust models to be trained. These models are able to provide plausible predictions of tissue types and allow accurate and objective tissue proportions to be computed. This will help to improve objectivity in downstream tasks such as pressure ulcer staging, healing risk prediction, identification on nonhealing wounds, adjustment of treatment options and may ultimately lead to improved healing rates for chronic wounds.

Acknowledgments

The authors would like to thank Dr Bob Bartlett and Dr Sheila Wang for their helpful suggestions in the preparation of this manuscript.

Conflicts of Interest

DR, JLRG, RDJF, and JA are current employees of Swift Medical Inc. At the time this research was conducted, JLRG was attached to the Division of Experimental Surgery, Department of Surgery, McGill University, Quebec, and RDJF was an adjunct Professor at the Arthur Labatt Family School of Nursing, Western University, Ontario.

Multimedia Appendix 1
Overview of research approach and additional results.
[PDF File (Adobe PDF File), 8233 KB - mhealth_v10i4e36977_app1.pdf]

Multimedia Appendix 2
Screen capture of our mobile app implementation of wound tissue segmentation.
[PDF File (Adobe PDF File), 310 KB - mhealth_v10i4e36977_app2.pdf]

References

1. Las Heras K, Igartua M, Santos-Vizcaino E, Hernandez RM. Chronic wounds: current status, available strategies and emerging therapeutic solutions. J Control Release 2020 Dec 10;328:532-550. [doi: 10.1016/j.jconrel.2020.09.039] [Medline: 32971198]
2. Sen CK, Gordillo GM, Roy S, Kirsner R, Lambert L, Hunt TK, et al. Human skin wounds: a major and snowballing threat to public health and the economy. Wound Repair Regen 2009;17(6):763-771 [FREE Full text] [doi: 10.1111/j.1524-475X.2009.00543.x] [Medline: 19903300]
3. Hopkins RB, Burke N, Harlock J, Jegathisawaran J, Goeree R. Economic burden of illness associated with diabetic foot ulcers in Canada. BMC Health Serv Res 2015 Jan 22;15:13 [FREE Full text] [doi: 10.1186/s12913-015-0687-5] [Medline: 25608648]
4. Stremitzer S, Wild T, Hoelzenbein T. How precise is the evaluation of chronic wounds by health care professionals? Int Wound J 2007 Jun;4(2):156-161 [FREE Full text] [doi: 10.1111/j.1742-481X.2007.00334.x] [Medline: 17651230]
5. Bates-Jensen BM. Chronic wound assessment. Nurs Clin North Am 1999 Dec;34(4):799-845. [Medline: 10523437]
6. Edsberg LE, Black JM, Goldberg M, McNichol L, Moore L, Sieggreen M. Revised national pressure ulcer advisory panel pressure injury staging system: revised pressure injury staging system. J Wound Ostomy Continence Nurs 2016;43(6):585-597 [FREE Full text] [doi: 10.1097/WON.0000000000000281] [Medline: 27749790]
7. Khoo R, Jansen S. The evolving field of wound measurement techniques: a literature review. Wounds 2016 Jun;28(6):175-181 [FREE Full text] [Medline: 27377609]
8. Shah A, Wollak C, Shah JB. Wound measurement techniques: comparing the use of ruler method, 2D imaging and 3D scanner. J Am Coll Clin Wound Spec 2015 Feb 27;5(3):52-57 [FREE Full text] [doi: 10.1016/j.jccw.2015.02.001] [Medline: 26199893]
9. Kass M, Witkin A, Terzopoulos D. Snakes: active contour models. Int J Comput Vision 1988 Jan;1(4):321-331. [doi: 10.1007/bf00133570]
10. Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts. IEEE Trans Pattern Anal Machine Intell 2001 Nov;23(11):1222-1239. [doi: 10.1109/34.969114]
11. Kolesnik M, Fexa A. Multi-dimensional color histograms for segmentation of wounds in images. In: Proceedings of the 2nd International Conference Image Analysis and Recognition. 2005 Presented at: ICIAR '05; September 28-30, 2005; Toronto, Canada p. 1014-1022. [doi: 10.1007/11559573_123]
12. Kolesnik M, Fexa A. How robust is the SVM wound segmentation? In: Proceedings of the 7th Nordic Signal Processing Symposium. 2006 Presented at: NORSIG '06; June 7-9, 2006; Reykjavik, Iceland p. 50-53. [doi: 10.1109/norsig.2006.275274]
13. Wang L, Pedersen PC, Agu E, Strong DM, Tulu B. Area determination of diabetic foot ulcer images using a cascaded two-stage SVM-based classification. IEEE Trans Biomed Eng 2017 Sep;64(9):2098-2109. [doi: 10.1109/TBME.2016.2632522] [Medline: 27893380]
14. Song B, Sacan A. Automated wound identification system based on image segmentation and artificial neural networks. In: IEEE International Conference on Bioinformatics and Biomedicine. 2012 Presented at: BIBM '12; October 4-7, 2012; Philadelphia, PA, USA p. 1-4. [doi: 10.1109/bibm.2012.6392633]
15. Wang C, Yan X, Smith M, Kochhar K, Rubin M, Warren SM, et al. A unified framework for automatic wound segmentation and analysis with deep convolutional neural networks. Annu Int Conf IEEE Eng Med Biol Soc 2015;2015:2415-2418. [doi: 10.1109/EMBC.2015.7318881] [Medline: 26736781]
16. Lu H, Li B, Zhu J, Li Y, Li Y, Xu X, et al. Wound intensity correction and segmentation with convolutional neural networks. Concurr Comput 2017 Mar 25;29(6):e3927. [doi: 10.1002/cpe.3927]
17. Goyal M, Yap MH, Reeves ND, Rajbhandari S, Spragg J. Fully convolutional networks for diabetic foot ulcer segmentation. In: IEEE International Conference on Systems, Man, and Cybernetics. 2017 Presented at: SMC '17; October 5-8, 2017; Banff, AB, Canada p. 618-623. [doi: 10.1109/smc.2017.8122675]
18. Yadav MK, Manohar DD, Mukherjee G, Chakraborty C. Segmentation of chronic wound areas by clustering techniques using selected color space. J Med Imaging Health Inform 2013 Mar 01;3(1):22-29. [doi: 10.1166/jmihi.2013.1124]
19. Li F, Wang C, Liu X, Peng Y, Jin S. A composite model of wound segmentation based on traditional methods and deep neural networks. Comput Intell Neurosci 2018 May 31;2018:4149103 [FREE Full text] [doi: 10.1155/2018/4149103] [Medline: 29955227]
20. Chakraborty C. Computational approach for chronic wound tissue characterization. Inform Med Unlocked 2019;17:100162. [doi: 10.1016/j.imu.2019.100162]
21. Wang C, Anisuzzaman DM, Williamson V, Dhar MK, Rostami B, Niezgoda J, et al. Fully automatic wound segmentation with deep convolutional neural networks. Sci Rep 2020 Dec 14;10(1):21897 [FREE Full text] [doi: 10.1038/s41598-020-78799-w] [Medline: 33318503]
22. Scebba G, Zhang J, Catanzaro S, Mihai C, Distler O, Berli M, et al. Detect-and-segment: a deep learning approach to automate wound image segmentation. Inform Med Unlocked 2022;29:100884. [doi: 10.1016/j.imu.2022.100884]
23. Perez AA, Gonzaga A, Alves JM. Segmentation and analysis of leg ulcers color images. In: Proceedings of the International Workshop on Medical Imaging and Augmented Reality. 2001 Presented at: MIAR '01; June 10-12, 2001; Hong Kong, China p. 262-266. [doi: 10.1109/miar.2001.930300]

24. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. 2015 Presented at: CVPR '15; June 7-12, 2015; Boston, MA, USA p. 3431-3440. [doi: 10.1109/cvpr.2015.7298965]
25. Liljequist D, Elfving B, Skavberg Roaldsen K. Intraclass correlation - a discussion and demonstration of basic features. PLoS One 2019 Jul 22;14(7):e0219854 [FREE Full text] [doi: 10.1371/journal.pone.0219854] [Medline: 31329615]
26. Computing Krippendorff 's alpha-reliability. University of Pennsylvania. 2011. URL: http://repository.upenn.edu/asc_papers/43 [accessed 2022-04-02]
27. Redmon J, Farhadi A. Yolov3: an incremental improvement. arXiv. Preprint posted online April 8, 2018 [FREE Full text]
28. Schlemper J, Oktay O, Schaap M, Heinrich M, Kainz B, Glocker B, et al. Attention gated networks: learning to leverage salient regions in medical images. Med Image Anal 2019 Apr;53:197-207 [FREE Full text] [doi: 10.1016/j.media.2019.01.012] [Medline: 30802813]
29. Chollet F. Xception: deep learning with depthwise separable convolutions. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. 2017 Presented at: CVPR '17; July 21-26, 2017; Honolulu, HI p. 1800-1807. [doi: 10.1109/cvpr.2017.195]
30. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 2014;15(56):1929-1958.
31. Tan M, Le Q. Efficientnet: rethinking model scaling for convolutional neural networks. In: Proceedings of the 36th International Conference on Machine Learning. 2019 Presented at: PMLR '19; June 9-15, 2019; Long Beach, CA, USA p. 6105-6114.
32. DeVries T, Taylor GW. Improved regularization of convolutional neural networks with cutout. arXiv. Preprint posted online August 15, 2017 [FREE Full text]
33. Loshchilov I, Hutter F. Decoupled weight decay regularization. arXiv. Preprint posted online November 14, 2017 [FREE Full text]
34. McGill University Dataverse. Scholars Portal Dataverse. 2022. URL: https://dataverse.scholarsportal.info/dataverse/jrg_experimental_surgery [accessed 2022-04-13]
35. Arpit D, Jastrzębski S, Ballas N, Krueger D, Bengio J, Kanwal MS, et al. A closer look at memorization in deep networks. In: Proceedings of the 34th International Conference on Machine Learning. 2017 Presented at: ICML '17; August 6-11, 2017; Sydney, Australia p. 233-242.
36. Matsoukas C, Hernandez AB, Liu Y, Dembrower K, Miranda G, Konuk E, et al. Adding seemingly uninformative labels helps in low data regimes. In: Proceedings of the 37th International Conference on Machine Learning. 2020 Presented at: ICML '20; July 13-18, 2020; Virtual p. 6775-6784.
37. Maier-Hein L, Mersmann S, Kondermann D, Bodenstedt S, Sanchez A, Stock C, et al. Can masses of non-experts train highly accurate image classifiers? A crowdsourcing approach to instrument segmentation in laparoscopic images. Med Image Comput Comput Assist Interv 2014;17(Pt 2):438-445. [doi: 10.1007/978-3-319-10470-6_55] [Medline: 25485409]
38. Heim E, Roß T, Seitel A, März K, Stieltjes B, Eisenmann M, et al. Large-scale medical image annotation with crowd-powered algorithms. J Med Imaging (Bellingham) 2018 Jul;5(3):034002 [FREE Full text] [doi: 10.1117/1.JMI.5.3.034002] [Medline: 30840724]
39. Karimi D, Dou H, Warfield SK, Gholipour A. Deep learning with noisy labels: exploring techniques and remedies in medical image analysis. Med Image Anal 2020 Oct;65:101759 [FREE Full text] [doi: 10.1016/j.media.2020.101759] [Medline: 32623277]

Abbreviations

BWAT: Bates-Jensen Wound Assessment Tool
ICC: intraclass correlation
mIOU: mean intersection over union
PUSH: Pressure Ulcer Scale for Healing

Edited by L Buis; submitted 01.02.22; peer-reviewed by S Shams, JA Benítez-Andrades, B Puladi; comments to author 24.02.22; revised version received 14.03.22; accepted 21.03.22; published 22.04.22.

Please cite as:
Ramachandram D, Ramirez-GarciaLuna JL, Fraser RDJ, Martínez-Jiménez MA, Arriaga-Caballero JE, Allport J
Fully Automated Wound Tissue Segmentation Using Deep Learning on Mobile Devices: Cohort Study
JMIR Mhealth Uhealth 2022;10(4):e36977
URL: https://mhealth.jmir.org/2022/4/e36977
doi:10.2196/36977
PMID:35451982

Original Paper

Digital Self-monitoring of Multiple Sclerosis: Interview Study With Dutch Health Care Providers on the Expected New Configuration of Roles and Responsibilities

Karine Wendrich[1], MSc; Lotte Krabbenborg[1], PhD

Institute for Science in Society, Radboud University Nijmegen, Nijmegen, Netherlands

Corresponding Author:
Lotte Krabbenborg, PhD
Institute for Science in Society
Radboud University Nijmegen
Heyendaalseweg 135
Nijmegen, 6525AJ
Netherlands
Phone: 31 024 3652269
Email: Lotte.Krabbenborg@ru.nl

Abstract

Background: Digital self-monitoring allows patients to produce and share personal health data collected at home. This creates a novel situation in which health care providers and patients must engage in a reconfiguration of roles and responsibilities. Although existing research pays considerable attention to the perceptions of patients regarding digital self-monitoring, less attention has been paid to the needs, wishes, and concerns of health care providers. As several companies and public institutions are developing and testing digital self-monitoring at the time of writing, it is timely and relevant to explore how health care providers envision using these technologies in their daily work practices. Our findings can be considered in decision-making processes concerning the further development and implementation of digital self-monitoring.

Objective: This study aims to explore how health care providers envisage using smartphone apps for digital self-monitoring of multiple sclerosis (MS) in their daily work practices, with a particular focus on physician-patient communication and on how health care providers respond to self-monitoring data and delegate tasks and responsibilities to patients.

Methods: We conducted semistructured in-depth interviews with 14 MS health care providers: 4 neurologists, 7 MS specialist nurses, and 3 rehabilitation professionals. They are affiliated with 3 different hospitals in the Netherlands that will participate in a pilot study to assess the efficiency and effectiveness of a specific smartphone app for self-monitoring.

Results: The interviewed health care providers seemed willing to use these smartphone apps and valued the quantitative data they produce that can complement the narratives that patients provide during medical appointments. The health care providers primarily want to use digital self-monitoring via prescription, meaning that they want a standardized smartphone app and want to act as its gatekeepers. Furthermore, they envisioned delegating particular tasks and responsibilities to patients via digital self-monitoring, such as sharing data with the health care providers or acting on the data, if necessary. The health care providers expected patients to become more proactive in the management of their disease. However, they also acknowledged that not all patients are willing or able to use digital self-monitoring apps and were concerned about the potential psychological and emotional burden on patients caused by this technology.

Conclusions: Our findings show that health care providers envisage a particular type of patient empowerment and personalized health care in which tensions arise between health care providers acting as gatekeepers and patient autonomy, between patient empowerment and patient disempowerment, and between the weight given to quantitative objective data and that given to patients' subjective experiences. In future research, it would be very interesting to investigate the actual experiences of health care providers with regard to digital self-monitoring to ascertain how the tensions mentioned in this paper play out in practice.

(JMIR Mhealth Uhealth 2022;10(4):e30224) doi:10.2196/30224

KEYWORDS

digital self-monitoring; smartphone apps; multiple sclerosis; technology assessment; health care providers; user participation; mobile phone

Introduction

Background

Increasing attention is being paid to digital self-monitoring by patients, that is, patients using digital devices to collect and record personal health data on bodily functions and everyday activities, such as various physical activities, mental status, and sleep patterns [1,2]. Smartphone apps, which are software programs designed for mobile devices, can be used for digital self-monitoring [3,4]. These apps make it relatively easy to produce and share personal health data and can therefore facilitate the maintenance of health and the self-management of chronic conditions [4]. Digital self-monitoring has implications for current care practices, as the setting for care is shifting from the hospital to the home, leading to new roles and responsibilities for health care providers as well as for patients [5,6]. There is considerable interest in using digital self-monitoring technologies to facilitate monitoring and self-management of multiple sclerosis (MS), a chronic neurological disease [7,8]. MS is characterized by a high variability of unpredictable symptoms, and the disease course is also unpredictable. Common symptoms are fatigue, mobility issues, and cognitive problems [9,10]. Digital self-monitoring is thought to support people with MS to self-manage their health in their home setting, for instance, by monitoring symptoms and adapting lifestyle behaviors [7,11-13].

Although existing research pays considerable attention to the motivations and perceptions of patients regarding digital self-monitoring, less attention has been paid to the needs, wishes, and concerns of health care providers [13,14]. Our research fills this gap and focuses particularly on how health care providers envision their new roles and responsibilities. Several scholars in the field of medical sociology (eg, the studies by Krabbenborg [15], Oudshoorn [16], Burri [17], and Pols [18]) have already shown how the introduction of new medical technologies in existing care infrastructures, be they eHealth technologies or otherwise, destabilizes the roles and responsibilities of health care providers. Burri [17], for example, showed how new digital imaging technologies challenged the identity and expertise of radiologists, as they had to acquire new knowledge and skills to interpret the images and use the machines [17]. Because their expertise was questioned by the introduction of these new technologies, radiologists were forced to re-establish their position as visual experts within the medical field. Pols [18] discussed how telecare technologies changed the daily working practices of nurses and their notions of good care [18]. Although nurses are used to seeing patients face-to-face and feared that telecare technologies might challenge their relationships with patients, they found that these technologies actually supported them in detecting patients' problems in a timely manner.

Similar processes might occur when MS health care providers are confronted with digital self-monitoring technologies. First, because care is shifting from traditional health care settings to patients' homes, health care providers might have to delegate some of their tasks and responsibilities to patients, such as recording measurements and interpreting data [5]. In addition, the health care provider becomes a coach who helps patients achieve their disease management goals via self-monitoring tools [14]. Second, self-monitoring can help patients gain, in principle, a better understanding of their condition. These insights can be shared with health care providers and could support their work. Although physician-patient communication could also benefit from the use of the technologies, disagreements could arise concerning whether the expertise of patients or the expertise of health care providers is the most valued [14]. Third, health care providers will have to learn how to deal with and respond to the continuous flow of data produced by digital self-monitoring [19]. The interpretation of these data can be difficult for health care providers, as they have to cope with variability in the data and decide when and how to take action in response to the data [20].

Objectives

At the time of writing, digital self-monitoring is not common practice in MS health care. However, as several companies and public institutions, such as universities, are developing and testing smartphone apps for self-monitoring MS [7], it is appropriate to explore how health care providers envisage using these technologies in their daily work practices. To this end, we conducted interviews with neurologists, MS specialist nurses, and rehabilitation specialists in the Netherlands. These 3 groups of health care providers are the most important in the MS care ecosystem. Neurologists mainly focus on the medical aspects of the disease, such as monitoring the progress of the disease and the effectiveness of medication. MS specialist nurses have a broader perspective; they also deal with psychosocial matters, such as how the disease affects patients' everyday lives, and are the connection between the various disciplines within the MS health care system. Most patients have appointments with their neurologist and MS specialist nurse approximately once or twice a year. Rehabilitation specialists, such as rehabilitation physicians and occupational therapists, are only visited when patients are finding it difficult to do their usual daily activities. The results of our investigations into the expectations of these health care providers can be considered in the further development and implementation of digital self-monitoring in health care.

Methods

Data Collection

We interviewed 14 MS health care providers: 4 neurologists, 7 MS specialist nurses, and 3 rehabilitation specialists (2 rehabilitation physicians and 1 occupational therapist) working at 3 different hospitals in the Netherlands, 1 academic hospital and 2 peripheral hospitals (Table 1). Of the 14 health care providers, 2 do not work at a hospital; they work for a care organization that was collaborating with one of the peripheral hospitals at the time of our study. The interviewed health care providers are all MS specialists, meaning that next to their general medical education they have been trained for treating patients with MS specifically.

Table 1. Overview of interviewed MS[a] health care providers, including their function and the hospital (or associated care organization) they work for.

Health care provider	Function	Hospital
MS1n[b]	Neurologist	Hospital 1 (academic hospital)
MS2n	Neurologist	Hospital 1 (academic hospital)
MS3sn[c]	MS specialist nurse	Hospital 1 (academic hospital)
MS4sn	MS specialist nurse	Hospital 1 (academic hospital)
MS5r[d]	Rehabilitation physician	Hospital 1 (academic hospital)
MS6n	Neurologist	Hospital 2 (peripheral hospital)
MS7sn	MS specialist nurse	Hospital 2 (peripheral hospital)
MS8sn	MS specialist nurse	Hospital 2 (peripheral hospital)
MS9sn	MS specialist nurse	Care organization collaborating with hospital 2
MS10r	Occupational therapist	Hospital 2 (peripheral hospital)
MS11r	Rehabilitation physician	Care organization collaborating with hospital 2
MS12n	Neurologist	Hospital 3 (peripheral hospital)
MS13sn	MS specialist nurse	Hospital 3 (peripheral hospital)
MS14sn	MS specialist nurse	Hospital 3 (peripheral hospital)

[a]MS: multiple sclerosis.

[b]MSn: neurologist.

[c]MSsn: specialist nurse.

[d]MSr: rehabilitation specialist (rehabilitation physician or occupational therapist).

The interviewed health care providers were purposefully sampled and are affiliated with hospitals that at the time of the interview were planning to participate in a pilot project to test the efficiency and effectiveness of a specific smartphone app for digitally self-monitoring MS, namely, the MS Sherpa app [21]. This app contains several weekly tests, such as a walking test and a cognition test, and daily questions, about, for example, mood, energy, and stress; it aims to enable patients with MS and their health care providers to monitor patients' mental and physical well-being. Before the interviews, we did not have any form of formal or informal collaboration with the interview respondents.

The health care providers were approached by mail by the first author (KW). For each hospital, we approached a neurologist, who subsequently connected us to the other MS health care providers working at their respective hospital. Of the 19 approached health care providers, 5 did not respond or did not want to participate in the interview. As data saturation occurred after the 14 health care providers were interviewed, meaning that no new topics or perspectives emerged, we decided not to approach additional health care providers.

The interviews were conducted between March and June 2019 and were performed by the first author KW, a PhD researcher in science and technology studies who has prior experience in conducting interviews with patients with MS and health care providers. The health care providers, who have busy schedules, could make an appointment for an interview either in person or by telephone, so as to enable some flexibility. In the end, a total of 14 interviews were conducted, 7 in person and 7 by telephone. The interviews lasted from 34 to 72 minutes, with an average duration of 56 (SD 10.08) minutes. The interview respondents received a summary of the main interview findings with the invitation to provide feedback on any false statements or other remarks, but no comments were returned.

A semistructured interview protocol was developed by the first author (KW) and the second author (LK), who is an associate professor in science and technology studies and experienced in conducting interviews with patients and health care providers about the possible effects of new biomedical innovations for their daily (work) life. The interview guide was inspired by the theme of the changing roles and responsibilities caused by digital self-monitoring, as mentioned in the *Introduction* section. The questions aim to gain an understanding of how health care providers envision digital self-monitoring of MS through smartphone apps to affect their daily working practices. Examples of interview topics were physician-patient communication, acting on self-monitoring data as a health care provider and delegating tasks and responsibilities to patients. The same interview guide was used for all interviews, but the semistructured setup of our interview guide allowed for some flexibility, for instance, to ask probing, open-ended questions on topics mentioned in previous interviews.

Ethics Approval

The interviews were audiotaped after verbal informed consent of the respondents, which has been approved by the Research Ethics Committee of the Faculty of Science (REC19012).

Data Analysis

First, the interviews were transcribed verbatim by the first author (KW) in Microsoft Word (2016; Microsoft Corporation). Next,

both authors familiarized themselves with the data by reading the transcripts a couple of times. The transcripts were then uploaded to the qualitative data analysis software ATLAS.ti 8 (2016; ATLAS.ti Scientific Software Development GmbH) and subjected to a combination of deductive and inductive thematic content analysis [22], meaning that we aimed for systematically structuring the interview data according to the main themes and subthemes. An initial codebook was developed by the first author (KW) based on discussions with the second author (LK) and was structured according to the topics in the interview guide. The codebook was refined through iterative reading of the interview transcripts, which resulted in additional topics being added to the codebook. Final agreement on the codebook was achieved through discussion between both authors. After coding the interview transcripts, the first author (KW) started clustering the codes into subthemes, which were subsequently clustered into main themes (Textbox 1). Consensus on themes and subthemes was reached through discussion with the second author (LK). Quotes translated from Dutch (by the first author) are used throughout the *Results* section to illustrate our findings.

Textbox 1. Overview of the main themes and subthemes that emerged from the interviews with the health care providers.

Perceived value of digital self-monitoring

- Providing additional and more complete information
- Quantifying patients' health status

Envisioned use of digital self-monitoring in daily working practices

- Digital self-monitoring on prescription
- Health care providers' access to digital self-monitoring data
- Preparation for the medical appointment
- Workload

Delegation of tasks and responsibilities to the patient

- More active role for patients in disease management
- Psychological and emotional burden of digital self-monitoring

Results

Perceived Value of Digital Self-monitoring

The interviewed health care providers expressed an ambivalent attitude toward digital self-monitoring, not only acknowledging its potential value but also expressing doubts and concerns. The longitudinal, real-world data generated by the use of smartphone apps might provide MS health care providers with information about the clinical monitoring of symptoms, disease progression, response to treatment, and side effects of medication, thereby supplementing the information collected during the medical appointment [7,8]. The interviewed health care providers indeed perceived the value of the information gained from digital self-monitoring, as they felt that information was often missing about the patient because appointments are short and there is insufficient time to perform all physical assessments, such as walking tests. Moreover, patients only have appointments once or twice a year, so it can be difficult for them to recall how their health has been over the preceding months. The health care providers believed that digital self-monitoring would produce additional and more complete information about patients' health status. Interestingly, the health care providers perceived that the quantitative data that would be produced by digital self-monitoring would be more objective than the narratives that patients tell during an appointment:

> *It would provide a more objective representation of something instead of what a patient is indicating subjectively.* [MS10r]

There seemed to be a desire to quantify a patient's health status to make a better assessment to complement the sensory and perceptual experiences of patients:

> *What I could do, is provide a better quantification. Which is often very difficult to find out during a conversation. How much is it exactly? How far can you walk? How many steps can you take? That kind of stuff. So, the quantification of all these things that we discuss should be the added value of a digital system.* [MS5r]

However, 2 health care providers emphasized that when there is too much of a focus on quantitative data, one runs the risk of ignoring the broader disease context, such as the psychosocial impact of MS:

> *If we are only occupied with measures and numbers. Will we not miss what the disease means to the patient? Not only how many meters they can walk. Because that it not always what they find the worst. But also talking about the psychosocial effect. I fear that the app misses that. That part should not be forgotten.* [MS1n]

Envisioned Use of Digital Self-monitoring in Daily Working Practices

Digital Self-monitoring on Prescription

There are various ways in which patients can gain access to a smartphone app to self-monitor their health: they can download an app that they have found themselves and that is freely

available on the commercial market, or they can use an app that has been recommended by health care providers and prescribed to patients as part of their treatment. The interviewed health care providers seemed to prefer digital self-monitoring on prescription, so that they can control which app patients use. They explained that if patients find an app themselves on the commercial market and therefore different patients use different apps, it will be difficult for health care providers to interpret the data and assess the reliability of each app. In contrast, when hospitals offer the same app that has been clinically validated to all patients, health care providers can become, over time, knowledgeable about a specific app. For instance, they can make comparisons between different patients:

> *I prefer to offer everyone the same app. Because if everyone uses something different, well, then it is just difficult to keep track of it. How reliable it is what they are doing. If something has been standardized and validated for what you are measuring. Then you can indeed offer it as a standard to patients so that you also get used to the numbers.* [MS1n]

This quote suggests that health care providers like to stay in charge when it comes to the digital self-monitoring of MS. They stressed that offering a smartphone app to patients should be an optional part of hospital treatment and only considered when the health care provider believes it could be valuable for the patient and when the patient is interested. As the health care providers explained, some patients are enthusiastic about collecting their personal health data, but others are not. Therefore, multiple health care providers emphasized that they have to consider whether digital self-monitoring is suitable for a specific patient. Health problems, such as cognitive or physical dysfunctions or limited (digital) health literacy, might hinder the use of digital self-monitoring by patients with MS. When health care providers are deciding which patients will be prescribed an app, they are at risk of becoming gatekeepers who control patients' access to digital self-monitoring.

Health Care Providers' Access to Digital Self-monitoring Data

Patients' digital self-monitoring raises a question about who should be able to access the collected data: the patient, the health care provider, or both. A common belief among the interviewed health care providers was that patients should control who has access to these data. All respondents agreed that these data should be accessible to patients and that the idea that patients cannot view their own health data is old-fashioned:

> *Well, these are the patient's data. So at the current time I do not think you can say that these are not available to them.* [MS2n]

The health care providers believed that they should only have access to the self-monitoring data after the patient had consented to this access. They envisioned that patients could select the data they wanted to be accessible to a particular health care provider. Most of the health care providers thought that it would be valuable to have access to the self-monitoring data. However, 2 health care providers expressed some doubts about patients' data becoming available to them, as this quote illustrates:

> *That would be sort of a Big Brother—that I can see from my computer how all my patients are doing. I am a bit hesitant about that.* [MS5r]

Having easy access to the self-monitoring data was crucial to the health care providers, as they have limited time to prepare for each appointment. The health care providers thought that a significant barrier is created when access to self-monitoring data is not straightforward, for example, because it involves logging in on a different system. As a respondent explained, their hospital had stopped using a web-based portal for digital self-monitoring because it took too much effort to access the data:

> *We have used [name of web-based portal] for a while. But it was just not workable. It was all external. And then we have to open it next to our file and that is just really difficult. That we have to go to their site to access patients' data. Well, if I have to do this in my half hour slot, I will not make it.* [MS7sn]

All health care providers preferred the digital self-monitoring data to be integrated into the patient's electronic file, because this made accessing the data easy. However, they also acknowledged that this would be technically complicated, as systems are often not compatible with each other.

Preparation for the Medical Appointment

The health care providers expected that they would mainly use the digital self-monitoring data to prepare for an appointment with a patient, so that they could focus on the patient's most prominent problems:

> *I think you can make your conversation more to the point. That you can leave out some domains where there are no problems. So I can use the remaining conversation time to focus on the areas where there are problems and that the patient has questions about. So I think it results in more directed conversations during the medical appointment.* [MS5r]

Although previous research has shown that patients feel reassured when health care providers monitor their health data in between appointments [23,24], our findings show that, because of a lack of time, the health care providers did not expect that they would look at the self-monitoring data in between appointments, unless the app or the patient indicated that something may be wrong. This suggests that health care providers do not think that it is their responsibility to monitor these data to detect deviations from the expected results or deterioration in the condition.

Moreover, the health care providers thought that patients could use smartphone apps to indicate the topics they wanted to talk about, but this would require patients to be willing and able to take the lead when engaging in digital self-monitoring. As one of the respondents explained, this would allow an appointment to be adapted to an individual patient's needs:

> *Tomorrow you have a check-up with the rehabilitation physician. What would you like to discuss? If the patient submits that, I can prepare better. Now I notice that patients are saying: well, I wanted to*

discuss this and this, but there was no time. Such an app could help with that. [MS11r]

Workload

The interviewed health care providers were ambivalent about the impact that digital self-monitoring might have on their workload. On the one hand, they imagined that it could result in a more efficient workflow as the data could make the appointment with the patient more efficient. On the other hand, they stated that digital self-monitoring requires a new way of working, such as integrating the data into daily workflows. They thought that this would take additional time, especially at the beginning. Moreover, some health care providers were worried that digital self-monitoring might result in information overload.

Be careful that you do not get so much information that you drown in information that is not useful. Not so much information that you lose the overview completely. You have to be careful that it does not become overkill. [MS6n]

Delegation of Tasks and Responsibilities to Patients

Digital self-monitoring requires that health care providers, to a certain extent, take on the role of *coach at a distance*, as tasks such as collecting and interpreting data are delegated to the patient. Patients are expected to become proactive in the management of their own medical care and lifestyle. However, the health care providers indicated that much responsibility is already given to the patient in the management of MS and that patients with MS tend to take on an active role:

I think that among the patient groups, MS patients already have and take quite an active role. My experience is that most patients are quite willing to take charge. [MS6n]

The health care providers believed that digital self-monitoring could help patients with MS to have even more control. They imagined that it would enable patients with MS to follow the progress of their disease, find patterns in their personal data, and compare these data with their personal experiences:

What could be an aim, is that through time they can see what is happening with them: days that they are feeling better or worse; what influence this has on the things that they are tracking. [MS1n]

I think it can provide insight to patients about when symptoms are occurring and what kind of symptoms. Maybe it becomes easier to make connections between certain things. [MS3sn]

According to some health care providers, patients could act on these insights, for instance, by contacting the health care provider if necessary. They also mentioned how smartphone apps could support patients in their particular lifestyle, such as stimulating them to be physically active:

Well, I think that the motivation for a good lifestyle, for instance the motivation to increase movement, that it works for many people. That thing [the smartphone app] provides compliments. If you are the type for that and you are sensitive to that, then it works very well. [MS9sn]

Although the health care providers acknowledged the potential value of digital self-monitoring in terms of empowering patients with MS, they also expressed concerns regarding the reconfiguration of patients' roles and responsibilities. The health care providers referred to the potential psychological and emotional burden of digital self-monitoring. They explained that digital self-monitoring might confront patients with their disease and make them feel more like a patient. Multiple health care providers mentioned that because of their experiences with patients with MS, they know that many of them do not like to be preoccupied with their disease:

I think that also a lot of patients will say: I do not want to be confronted with my disease, so I am not going to use it. Right now I have some patients who are saying: I do not want to be confronted with my disease. I just want to live. [MS7sn]

Moreover, the health care providers were concerned that digital self-monitoring data might worry or disappoint patients, for instance, because there are signs that the disease is progressing. This suggests that health care providers are ambivalent about the reconfiguration of patients' roles and responsibilities.

Discussion

Principal Findings

At the time of writing, several smartphone apps that digitally self-monitor MS are being developed [7], which creates a novel situation in which health care providers and patients must engage in a reconfiguration of roles and responsibilities. From our interviews, we have gained insight into the expected new configuration from the perspective of MS health care providers. The health care providers in our study are generally willing to use smartphone apps for self-monitoring of patients and value the quantified data that become available because they complement patients' narratives that are revealed during medical appointments. We found that the health care providers primarily want to use digital self-monitoring on prescription, meaning that they want a standardized smartphone app that has been clinically validated. They also believed that digital self-monitoring should always be an option, rather than an obligation. They acknowledged that not all patients are willing or able to engage in digital self-monitoring and expressed concerns regarding the potential psychological and emotional burden of this technology. Furthermore, the health care providers envisioned a delegation of particular tasks and responsibilities to patients, such as sharing data with their health care provider and contacting the hospital if necessary. However, this new configuration would potentially bring with it considerable tensions and issues regarding the type of patient empowerment and personalized health care that the digital self-monitoring is aiming for, and we describe the main four next.

First, the prescription of self-monitoring apps by health care providers is in line with the idea of *clinical self-tracking* in which self-monitoring practices are pushed by health care providers, that is, self-monitoring by the patient for therapeutic purposes following the recommendation of health care providers [25]. In these clinical self-tracking practices, data are shared with health care providers, and data collection is standardized

to help health care providers analyze and interpret the data. However, if health care providers act as gatekeepers regarding patients' access to digital self-monitoring and standardized apps are used, this could restrict patients' autonomy and reinforce the *paternalistic model* in which health care providers rather than patients are in charge [26]. Will there be time during medical appointments for patients to discuss data produced by self-monitoring apps that they have purchased themselves on the commercial market, for example? Moreover, our previous research showed that patients with MS want their use of digital self-monitoring to be flexible, that is, to consider their personal situation and disease status [12]. However, because of what health care providers said in the interviews, we wonder whether patients will have the flexibility to deviate from self-monitoring protocols and whether, if they do so, health care providers might then distrust the data. Therefore, when self-monitoring apps are prescribed by health care providers, there might be a tension between the traditional paternalistic notion of physician-patient relationships and the ideal of patient empowerment, that is, patients being in charge of their own disease management, as promoted by the developers of these apps. When opting for a specific implementation strategy, such as self-monitoring on prescription, technology developers should consider that this implies a certain vision on health care.

Second, how realistic is it to think that patients will engage in the tasks and responsibilities as imagined by the health care providers in our study? Discrepancies can exist between the expectations of health care providers and the assumptions of patients regarding their roles and responsibilities. Although the health care providers in our study thought that patients would contact them if there were deviations in the self-monitoring data, previous research has found that patients are hesitant to do so as they downgrade their problems and do not want to bother their health care providers [27]. Furthermore, tensions can arise when patients and health care providers have different expectations about the amount of support that health care providers offer in the self-monitoring process [28]. Patients might feel reassured by a sense of being supervised by their health care providers [23,24], but the health care providers in our study stressed the personal responsibility of patients in the management of their disease, for instance, their responsibility for signaling potential signs of disease progression. The promises and expectations surrounding digital self-monitoring, as among others expressed by technology developers, are based on ideal situations and assume that patients have the right knowledge and skills to deal with the collected data and are willing and able to engage in self-managing their health [1,14]. However, as noted previously, patients' physical and cognitive capacities, and limited digital or health literacy, could pose a barrier to patients' abilities to adapt to the new roles and responsibilities required by digital self-monitoring, such as correctly interpreting their own data and reacting appropriately to it [29]. Moreover, digital self-monitoring could pose a psychological and emotional burden on patients, which might have a disempowering rather than an empowering effect on patients [1,2,30]. These factors imply that not all patients are willing or able to engage in digital self-monitoring, which might result in health inequities when self-monitoring apps become an integral part of health care [29]. In fact, as Prainsack [31] already argued, the introduction of technologies to personalize health care, such as self-monitoring apps, as well as genetic self-tests and home-based blood monitoring technologies, might confront us with fundamental questions such as what social circumstances and capacities do patients need to meet to participate in the shift toward more personalized, digital health care? In addition, which groups of patients tend to be included and excluded when digital self-monitoring gains more significance in health care?

Third, when digital self-monitoring reduces health to data and algorithms, the richness and complexity of patients' experiences can be lost [2]. Tensions might arise between the place that objective numbers will be given in health care and patients' subjective narratives [31]. The interviewed health care providers seemed to want quantitative assessments of patients' health status to complement patients' experiences. However, as 2 health care providers also acknowledged, this might result in too much of a focus on numbers while the broader context of the patient, such as their psychological struggles, is ignored [2]. When the conversation is dominated by the quantitative data during the medical appointment, patients have less time to share their personal stories. Consequently, the questions and issues that are the most pressing ones for the patient might not be discussed. Moreover, it is questionable how objective these quantitative data truly are. Although numbers seem objective and neutral, they are full of assumptions and value judgments, such as choices made regarding which parameters are being measured and how [1,2]. There might also be discrepancies between what the numbers are indicating and what the patients are experiencing. This brings us to the sociological concepts of *illness* and *disease*, with illness being defined as how patients identify with their ill-health and disease meaning the condition diagnosed by a health care provider, that is, as biologically defined [32]. When health care providers value quantitative self-monitoring data more than other factors, they may primarily focus on the biological condition, that is, the disease, whereas we know that illness is just as important for patients. These issues highlight that when integrating digital self-monitoring data into health care, one needs to consider how to balance quantitative data with patients' feelings and experiences when judging a patient's health status. This discussion is interesting in the light of the current paradigm shift toward *personalized health care* [2]. Our findings bring the following question to the fore: what makes personalized health care really personalized? Does personalized health care involve having more quantitative data on patients' health status, or is it about paying more attention to patients' personal experiences? Alternatively, should both elements be combined to achieve truly personalized health care? We recommend that the developers of digital self-monitoring technologies be aware of the fact that personalized health care is not a straightforward concept and can evoke different connotations for patients, health care providers, and technology developers.

Finally, although the health care providers in our study valued digital self-monitoring of MS, our findings also underline that is not self-evident that health care providers are willing and able to use this technology as part of their daily practices. The new roles and responsibilities demanded by the use of digital

self-monitoring, such as delegating the collection, interpretation, and sharing of data to patients, can be a challenge for health care providers and their professional identities and routines as our study shows. Moreover, as the literature on mobile health technologies for other disease domains has already highlighted, health care providers also experience other barriers, which are related to the quality, validity, accuracy, and clinical utility of mobile health technologies and patient-generated data [33-35]. As several studies have shown, health care providers have more confidence in data collected by themselves than by patients [30,34,36]. Furthermore, health care providers have raised medical legal concerns about data security, privacy, and accountability, such as the confidentiality of self-monitoring data and the responsibility of health care providers to act upon patient data [34,37].

These considerations illustrate the complexities surrounding the use of digital self-monitoring technologies by health care providers and also indicate that the embedding of these technologies in clinical practices cannot be decided by using a top-down approach. Instead, a mutual collaboration is needed between the technology developers and the health care providers who are expected to work with digital self-monitoring. There often seems to be a mismatch between the developers of mobile health technologies and the health care providers who ought to use these technologies, with insufficient consideration being given to user needs [37,38]. As Tarricone et al [37] argue, app developers should do more to include health care providers during the development process of health apps. When the developers of digital self-monitoring technologies adopt such an approach, implementation strategies for digital self-monitoring can be developed iteratively so that there is a greater chance that the technology fits the daily working practices of health care providers.

Study Limitations

As far as we know, this is the first qualitative study on the expectations, desires, and concerns of MS health care providers regarding digital self-monitoring. Although this study has provided a rich insight into the perspectives of health care providers, future studies would probably benefit from including more respondents per health care provider group (neurologists, specialist nurses, and rehabilitation specialists). This would allow for a more rigorous comparison between these groups, which goes beyond the aim of this study. In addition, other health care providers involved in the care of patients with MS, such as psychologists and physiotherapists, could be included in future studies.

Moreover, although we focused on MS specialists, as these respondents were thought to be the most informative for this study, not all patients with MS are treated by health care providers specializing in MS. In fact, many patients with MS are under treatment of general neurologists and rehabilitation specialists. It would be interesting to study the similarities and differences in needs, wishes, and experiences regarding digital self-monitoring of MS between the MS specialists and the general health care providers.

Another consideration we would like to point out is that the interviewed health care providers are affiliated with hospitals that will be enrolled in a pilot study on digital self-monitoring of MS. These health care providers might therefore have a biased positive attitude toward the value of digital self-monitoring compared with those who are not going to be engaged in the study. Finally, our study only investigated Dutch health care providers' attitudes. It might be interesting to compare our findings with the attitudes of health care providers in other countries because of cultural differences in different countries' health care systems.

Conclusions

To conclude, in this study, we have identified potential tensions and problems that might occur when digital self-monitoring is introduced in existing health care. Our findings could be considered in study protocols of future projects on digital self-monitoring or in implementation strategies for integrating these technologies into clinical practices. For instance, developers could consider what type of personalized health care and patient empowerment they want to stimulate with their technology, as this is not straightforward and patients and health care providers can have different visions in this regard. As there might be differences between what people say and the actual acts they perform, in future research, it would be very interesting to investigate the actual actions and experiences of health care providers with regard to digital self-monitoring to determine how the tensions mentioned in this paper are being played out in practice.

Acknowledgments

This work received financial support from the Dutch National MS Foundation and the Netherlands Organisation for Scientific Research (Data2Person Project 628.011.025). The authors would like to thank MS Sherpa BV for helping contact the health care providers.

Conflicts of Interest

The grant awarded by the Netherlands Organisation for Scientific Research required a cash contribution from a private party. MS Sherpa BV, a company that has developed a self-monitoring app for multiple sclerosis, provided a cash contribution of 15% of this subsidy. In line with our consortium agreement, our research was conducted independently. MS Sherpa BV had no influence on the design, analysis, or outcomes of our research.

References

1. Lupton D. Quantifying the body: monitoring and measuring health in the age of mHealth technologies. Crit Public Health 2013;23(4):393-403. [doi: 10.1080/09581596.2013.794931]
2. Sharon T. Self-tracking for health and the quantified self: re-articulating autonomy, solidarity, and authenticity in an age of personalized healthcare. Philos Technol 2017;30(1):93-121. [doi: 10.1007/s13347-016-0215-5]
3. Lupton D. Apps as artefacts: towards a critical perspective on mobile health and medical apps. Societies 2014 Oct 29;4(4):606-622. [doi: 10.3390/soc4040606]
4. Birkhoff SD, Smeltzer SC. Perceptions of smartphone user-centered mobile health tracking apps across various chronic illness populations: an integrative review. J Nurs Scholarsh 2017;49(4):371-378. [doi: 10.1111/jnu.12298] [Medline: 28605151]
5. Danesi G, Pralong M, Panese F, Burnand B, Grossen M. Techno-social reconfigurations in diabetes (self-) care. Soc Stud Sci 2020;50(2):198-220. [doi: 10.1177/0306312720903493] [Medline: 32009559]
6. Oudshoorn N. How places matter: telecare technologies and the changing spatial dimensions of healthcare. Soc Stud Sci 2012;42(1):121-142. [doi: 10.1177/0306312711431817] [Medline: 22530385]
7. Marziniak M, Brichetto G, Feys P, Meyding-Lamadé U, Vernon K, Meuth SG. The use of digital and remote communication technologies as a tool for multiple sclerosis management: narrative review. JMIR Rehabil Assist Technol 2018;5(1):e5 [FREE Full text] [doi: 10.2196/rehab.7805] [Medline: 29691208]
8. Simblett SK, Evans J, Greer B, Curtis H, Matcham F, Radaelli M, RADAR-CNS consortium. Engaging across dimensions of diversity: a cross-national perspective on mHealth tools for managing relapsing remitting and progressive multiple sclerosis. Mult Scler Relat Disord 2019;32:123-132. [doi: 10.1016/j.msard.2019.04.020] [Medline: 31125754]
9. Dobson R, Giovannoni G. Multiple sclerosis - a review. Eur J Neurol 2019;26(1):27-40. [doi: 10.1111/ene.13819] [Medline: 30300457]
10. Kamm CP, Uitdehaag BM, Polman CH. Multiple sclerosis: current knowledge and future outlook. Eur Neurol 2014;72(3-4):132-141 [FREE Full text] [doi: 10.1159/000360528] [Medline: 25095894]
11. Ayobi A, Marshall P, Cox AL. Quantifying the BodyCaring for the mind: self-tracking in multiple sclerosis. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI '17; May 6-11, 2017; Colorado, Denver, USA p. 6889-6901. [doi: 10.1145/3025453.3025869]
12. Wendrich K, van Oirschot P, Martens MB, Heerings M, Jongen PJ, Krabbenborg L. Toward digital self-monitoring of multiple sclerosis: investigating first experiences, needs, and wishes of people with MS. Int J MS Care 2019;21(6):282-291 [FREE Full text] [doi: 10.7224/1537-2073.2018-083] [Medline: 31889935]
13. Giunti G, Kool J, Rivera Romero O, Dorronzoro Zubiete E. Exploring the specific needs of persons with multiple sclerosis for mHealth solutions for physical activity: mixed-methods study. JMIR Mhealth Uhealth 2018;6(2):e37 [FREE Full text] [doi: 10.2196/mhealth.8996] [Medline: 29426814]
14. Gabriels K, Moerenhout T. Exploring entertainment medicine and professionalization of self-care: interview study among doctors on the potential effects of digital self-tracking. J Med Internet Res 2018;20(1):e10 [FREE Full text] [doi: 10.2196/jmir.8040] [Medline: 29330140]
15. Krabbenborg L. Dramatic rehearsal on the societal embedding of the lithium chip. In: van der Burg S, Swierstra T, editors. Ethics on the laboratory floor. London, UK: Palgrave Macmillan; 2013:168-187.
16. Oudshoorn N. Diagnosis at a distance: the invisible work of patients and healthcare professionals in cardiac telemonitoring technology. Sociol Health Illn 2008;30(2):272-288 [FREE Full text] [doi: 10.1111/j.1467-9566.2007.01032.x] [Medline: 18290936]
17. Burri RV. Doing distinctions: boundary work and symbolic capital in radiology. Soc Stud Sci 2008;38(1):35-62. [doi: 10.1177/0306312707082021]
18. Pols J. The heart of the matter: good nursing at a distance. In: Pols J, editor. Care at a distance: on the closeness of technology. Amsterdam, The Netherlands: Amsterdam University Press; 2012:45-60.
19. Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: promises and barriers. PLoS Med 2016;13(2):e1001953 [FREE Full text] [doi: 10.1371/journal.pmed.1001953] [Medline: 26836780]
20. Fletcher BR, Hinton L, Hartmann-Boyce J, Roberts NW, Bobrovitz N, McManus RJ. Self-monitoring blood pressure in hypertension, patient and provider perspectives: a systematic review and thematic synthesis. Patient Educ Couns 2016;99(2):210-219. [doi: 10.1016/j.pec.2015.08.026] [Medline: 26341941]
21. van Oirschot P, Heerings M, Wendrich K, den Teuling B, Martens MB, Jongen PJ. Symbol digit modalities test variant in a smartphone app for persons with multiple sclerosis: validation study. JMIR Mhealth Uhealth 2020;8(10):e18160 [FREE Full text] [doi: 10.2196/18160] [Medline: 33016886]
22. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: implications for conducting a qualitative descriptive study. Nurs Health Sci 2013;15(3):398-405. [doi: 10.1111/nhs.12048] [Medline: 23480423]
23. Ure J, Pinnock H, Hanley J, Kidd G, McCall Smith E, Tarling A, et al. Piloting tele-monitoring in COPD: a mixed methods exploration of issues in design and implementation. Prim Care Respir J 2012;21(1):57-64 [FREE Full text] [doi: 10.4104/pcrj.2011.00065] [Medline: 21785816]

24. Fairbrother P, Pinnock H, Hanley J, McCloughan L, Sheikh A, Pagliari C, et al. Exploring telemonitoring and self-management by patients with chronic obstructive pulmonary disease: a qualitative study embedded in a randomized controlled trial. Patient Educ Couns 2013;93(3):403-410. [doi: 10.1016/j.pec.2013.04.003] [Medline: 23647981]
25. Piras EM, Miele F. Clinical self-tracking and monitoring technologies: negotiations in the ICT-mediated patient–provider relationship. Health Sociol Rev 2016;26(1):38-53. [doi: 10.1080/14461242.2016.1212316]
26. Fernández-Ballesteros R, Sánchez-Izquierdo M, Olmos R, Huici C, Ribera Casado JM, Cruz Jentoft A. Paternalism vs. autonomy: are they alternative types of formal care? Front Psychol 2019;10:1460 [FREE Full text] [doi: 10.3389/fpsyg.2019.01460] [Medline: 31316428]
27. Vasileiou K, Barnett J, Young T. Interpreting and acting upon home blood pressure readings: a qualitative study. BMC Fam Pract 2013;14:97 [FREE Full text] [doi: 10.1186/1471-2296-14-97] [Medline: 23849055]
28. Irfan Khan A, Gill A, Cott C, Hans PK, Steele Gray C. mHealth tools for the self-management of patients with multimorbidity in primary care settings: pilot study to explore user experience. JMIR Mhealth Uhealth 2018;6(8):e171 [FREE Full text] [doi: 10.2196/mhealth.8593] [Medline: 30154073]
29. Bruins BJ. Gezondheidsapps en wearables De ethiek van e-health deel I. The Hague, The Netherlands: Centrum voor Ethiek en Gezondheid; 2020.
30. Ancker JS, Witteman HO, Hafeez B, Provencher T, Van de Graaf M, Wei E. "You get reminded you're a sick person": personal data tracking and patients with multiple chronic conditions. J Med Internet Res 2015;17(8):e202 [FREE Full text] [doi: 10.2196/jmir.4209] [Medline: 26290186]
31. Prainsack B. Personalized medicine: empowered patients in the 21st century. New York, New York, USA: New York University Press; 2017.
32. Hofmann B. Disease, illness and sickness. In: Solomon M, Simon JR, Kincaid H, editors. The Routledge companion to philosophy of medicine. London, UK: Routledge; 2016:16-26.
33. West P, Van Kleek M, Giordano R, Weal M, Shadbolt N. Information quality challenges of patient-generated data in clinical practice. Front Public Health 2017;5:284 [FREE Full text] [doi: 10.3389/fpubh.2017.00284] [Medline: 29209601]
34. Sarradon-Eck A, Bouchez T, Auroy L, Schuers M, Darmon D. Attitudes of general practitioners toward prescription of mobile health apps: qualitative study. JMIR Mhealth Uhealth 2021;9(3):e21795 [FREE Full text] [doi: 10.2196/21795] [Medline: 33661123]
35. Bellicha A, Macé S, Oppert J. Prescribing of electronic activity monitors in cardiometabolic diseases: qualitative interview-based study. J Med Internet Res 2017;19(9):e328 [FREE Full text] [doi: 10.2196/jmir.8107] [Medline: 28947415]
36. Meidert U, Scheermesser M. Quantified self-technologies for a better disease management? In: 2021 Congress of the Swiss Sociological Association. 2021 Presented at: SSA '21; June 28-30, 2021; Geneva, Switzerland.
37. Tarricone R, Cucciniello M, Armeni P, Petracca F, Desouza KC, Hall LK, et al. Mobile health divide between clinicians and patients in cancer care: results from a cross-sectional international survey. JMIR Mhealth Uhealth 2019;7(9):e13584 [FREE Full text] [doi: 10.2196/13584] [Medline: 31493318]
38. Rudin RS, Bates DW, MacRae C. Accelerating innovation in health IT. N Engl J Med 2016;375(9):815-817 [FREE Full text] [doi: 10.1056/NEJMp1606884] [Medline: 27579633]

Abbreviations

MS: multiple sclerosis

Edited by L Buis; submitted 06.05.21; peer-reviewed by O Rivera, J Long; comments to author 13.10.21; revised version received 04.11.21; accepted 18.02.22; published 27.04.22.

Please cite as:
Wendrich K, Krabbenborg L
Digital Self-monitoring of Multiple Sclerosis: Interview Study With Dutch Health Care Providers on the Expected New Configuration of Roles and Responsibilities
JMIR Mhealth Uhealth 2022;10(4):e30224
URL: https://mhealth.jmir.org/2022/4/e30224
doi:10.2196/30224
PMID:35475770

Original Paper

Residual Effect of Texting to Promote Medication Adherence for Villagers with Schizophrenia in China: 18-Month Follow-up Survey After the Randomized Controlled Trial Discontinuation

Yiyuan Cai[1*], PhD; Wenjie Gong[2,3,4*], PhD; Wenjun He[5,6], MM; Hua He[7], PhD; James P Hughes[8], PhD; Jane Simoni[9,10], PhD; Shuiyuan Xiao[11], PhD; Stephen Gloyd[10], PhD; Meijuan Lin[12], MM; Xinlei Deng[13], BA; Zichao Liang[13], BA; Bofeng Dai[13], MM; Jing Liao[13], PhD; Yuantao Hao[13], PhD; Dong Roman Xu[5,6], PhD

[1]Department of Epidemiology and Health Statistic, School of Public Health, Guizhou Medical University, Guiyang, China

[2]HER Team and Department of Maternal and Child Health, Xiangya School of Public Health, Central South University, Changsha, China

[3]Department of Psychiatry, University of Rochester Medical Center, Rochester, NY, United States

[4]Institute of Applied Health Research, University of Birmingham, Birmingham, United Kingdom

[5]Center for World Health Organization (WHO) Studies and Department of Health Management, School of Health Management of Southern Medical University, Guangzhou, China

[6]Acacia Labs, Institute for Global Health and Dermatology Hospital, Southern Medical University, Guangzhou, China

[7]School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA, United States

[8]Department of Biostatistics, University of Washington, Seattle, WA, United States

[9]Department of Psychology, University of Washington, Seattle, WA, United States

[10]Department of Global Health, University of Washington, Seattle, WA, United States

[11]Xiangya School of Public Health, Central South University, Changsha, China

[12]School of Nursing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong Special Administrative Region, China

[13]School of Public Health, Sun Yat-sen University, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Dong Roman Xu, PhD
Center for World Health Organization (WHO) Studies and Department of Health Management
School of Health Management of Southern Medical University
1023 South Shatai Road
Guangzhou, 510515
China
Phone: 86 02061647194
Email: romanxu@i.smu.edu.cn

Abstract

Background: Reducing the treatment gap for mental health in low- and middle-income countries is a high priority. Even with treatment, adherence to antipsychotics is rather low. Our integrated intervention package significantly improved medication adherence within 6 months for villagers with schizophrenia in resource-poor communities in rural China. However, considering the resource constraint, we need to test whether the effect of those behavior-shaping interventions may be maintained even after the suspension of the intervention.

Objective: The aim of this study is to explore the primary outcome of adherence and other outcomes at an 18-month follow-up after the intervention had been suspended.

Methods: In a 6-month randomized trial, 277 villagers with schizophrenia were randomized to receive either a government community mental health program (686 Program) or the 686 Program plus Lay health supporters, e-platform, award, and integration (LEAN), which included health supporters for medication or care supervision, e-platform access for sending mobile SMS text messaging reminders and education message, a token gift for positive behavior changes (eg, continuing taking medicine), and integrating the e-platform with the existing 686 Program. After the 6-month intervention, both groups received only the 686 Program for 18 months (phase 2). Outcomes at both phases included antipsychotic medication adherence, functioning, symptoms, number of rehospitalization, suicide, and violent behaviors. The adherence and functioning were assessed at the home visit by

trained assessors. We calculated the adherence in the past 30 days by counting the percentage of dosages taken from November to December 2018 by unannounced home-based pill counts. The functioning was assessed using the World Health Organization Disability Assessment Schedule 2.0. The symptoms were evaluated using the Clinical Global Impression–Schizophrenia during their visits to the 686 Program psychiatrists. Other outcomes were routinely collected in the 686 Program system. We used intention-to-treat analysis, and missing data were dealt with using multiple imputation. The generalized estimating equation model was used to assess program effects on adherence, functioning, and symptoms.

Results: In phase 1, antipsychotic adherence and rehospitalization incidence improved significantly. However, in phase 2, the difference of the mean of antipsychotic adherence (adjusted mean difference 0.05, 95% CI −0.06 to 0.16; *P*=.41; Cohen *d* effect size=0.11) and rehospitalization incidence (relative risk 0.65, 95% CI 0.32-1.33; *P*=.24; number needed to treat 21.83, 95% CI 8.30-34.69) was no longer statistically significant, and there was no improvement in other outcomes in either phase (*P*≥.05).

Conclusions: The simple community-based LEAN intervention could not continually improve adherence and reduce the rehospitalization of people with schizophrenia. Our study inclined to suggest that prompts for medication may be necessary to maintain medication adherence for people with schizophrenia, although we cannot definitively exclude other alternative interpretations.

(JMIR Mhealth Uhealth 2022;10(4):e33628) doi:10.2196/33628

KEYWORDS

medication adherence; mobile texting; lay health worker; resource-poor community; primary health care; quality of care; mHealth; schizophrenia; maintenance; residual effect; mental health; patient outcomes

Introduction

Background

Treatment with antipsychotic medication effectively prevents relapse and rehospitalization in people with schizophrenia [1-4]. However, 69% of people with schizophrenia in low- and middle-income countries have no access to any evidence-based care, often because of resource constraints or health system failures [5]. Even when treatment is available, adherence to antipsychotics is rather low [6], with nearly half of people with schizophrenia taking less than 70% of their prescribed doses [7]. In 2005, China launched the National Continuing Management and Intervention Program for Psychoses, also known as the 686 Program [8,9], which later became part of China's *integrated public mental health service*. The program followed the World Health Organization Mental Health Gap Action Program recommendations [10] and featured a collaborative approach between psychiatrists and community health workers to screen, diagnose, treat, and manage psychosis in communities [11,12]. In 2017, the 686 Program covered over 5,810,000 people with psychosis across China [11]. However, although the program provides free medication to the poor, over 70% of program enrollees failed to take their antipsychotic medications routinely [9].

China's 686 Program, which is likely the largest community-based mental health program in the world, benefited a vast population of people with schizophrenia. However, the program effect had been reduced because of the poor patient adherence to antipsychotics [6,9]. Schizophrenia can often be effectively controlled with lifelong antipsychotic medications [4]. Mobile texting or SMS text messaging was found to be an inexpensive, easy-to-use, and reliable means to improve treatment adherence and schizophrenia care in some high-income settings [13-15]. To test its effectiveness in the resource-poor setting, we conducted a pragmatic trial featuring SMS text messaging reminders as the core of intervention in 2016 in 9 rural communities of China (called the Lay health supporters, e-platform, award, and integration [LEAN] Trial) [6,16]. The 6-month trial showed that LEAN significantly improved medication adherence (proportion of dosages taken) from 0.48 to 0.61 (adjusted mean difference 0.11, 95% CI 0.03-0.20; *P*=.007) among villagers with schizophrenia, and the incidence of rehospitalization due to schizophrenia decreased from 19.5% (25/128) under the control arm to 7.3% (9/124) under the intervention arm (relative risk 0.32, 95% CI 0.18-0.76; *P*=.007). However, other outcomes, including patient functioning and symptoms measured in phase 1, did not present significant improvement.

Objectives

As of August 1, 2021, we have not yet identified other studies that have been explored the maintenance of the effect after a texting-based intervention was withdrawn in people with schizophrenia. Although we identified no similar studies that focused on the same participants as ours, studies of behavioral interventions using mobile texting in other populations suggested that the program effects lasted 8 to 12 weeks after the termination of texting [17]. Dobson et al [18] found that the impact of a texting intervention to improve self-management in patients with diabetes remained effective 15 months after the intervention was discontinued. Menon et al [19] found that the medication adherence in patients with bipolar I disorder persisted for at least three months after the intervention was discontinued. It is essential to understand the sustained effect of SMS text messaging after the intervention because (1) most interventions will not or cannot go on forever because of resource constraints or other reasons and (2) it may not be necessary to perpetuate a behavior-oriented intervention, as in theory, once people with schizophrenia habituate to new behavior, the original *reminder* may no longer be needed to stimulate the desired behavior. In the LEAN study, rural people with schizophrenia were randomized into either a mobile texting group (the LEAN intervention group) or a care-as-usual group (the control group) and were followed up for 6 months on LEAN's effect on medication adherence, symptoms, and

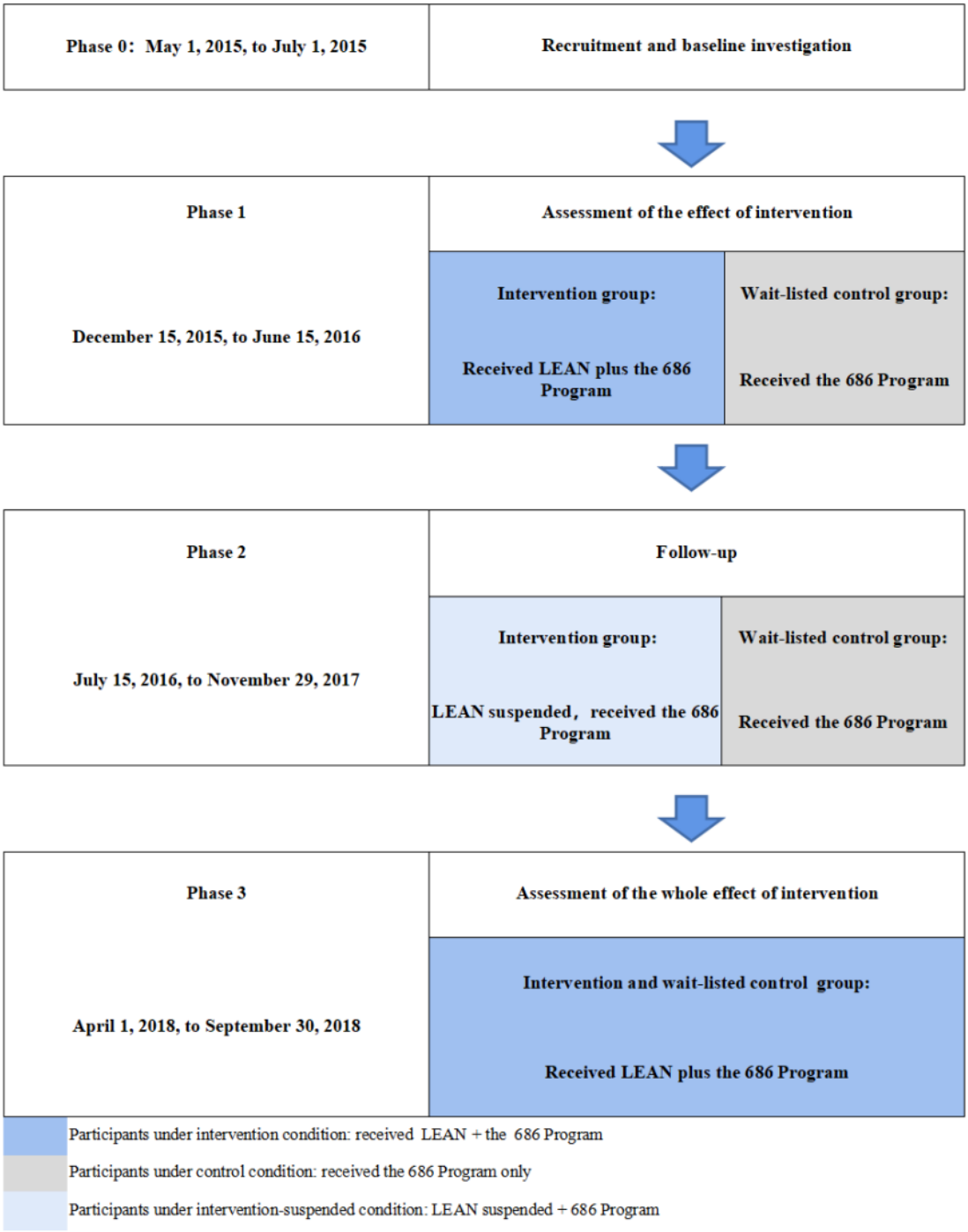
functioning. The LEAN intervention was then suspended, and we re-examined its sustained effect 18 months later. The study was divided into 3 phases. Phase 1 was a 2-arm randomized controlled study to evaluate the impact of LEAN in improving medication adherence. Phase 2 suspended the implementation of LEAN for the intervention group. Phase 3 restarted LEAN and applied it to both the original intervention and control groups and used information from all phases to evaluate the long-term effect of LEAN. Phase 1 and phase 3 results were published elsewhere [6,16], which established and reinforced the conclusion about LEAN's impact on improving medication adherence. This paper reported the results of phase 2. The suspended implementation of LEAN during this phase provided a unique opportunity to understand whether texting can establish a new medication-taking behavior that may make texting no longer necessary. We hypothesized that the effects would be sustained as people may have habituated to the new behavior in the initial 6-month intervention so that the reminders may no longer be necessary. Furthermore, we also hypothesize that, with the improvement of the symptoms, functioning, and other outcomes would also present improvements in the long term.

Methods

Trial Design

The whole program process was divided into 3 phases [20]. We conducted a 2-arm randomized controlled trial in phase 1. In the 6-month phase 1 (December 15, 2015, to June 15, 2016), the intervention group received LEAN plus the aforementioned 686 Program for 6 months and the control group received the 686 Program only (trial registration: Chinese Clinical Trial Registry ChiCTR-ICR-15006053) [16]. After that, the LEAN intervention was suspended, and both groups received the 686 Program only (phase 2: from July 15, 2016, to November 29, 2017). Then, we restarted our intervention and extended it to both the original intervention group and the control group, and all participants received the LEAN intervention in addition to the 686 Program from April 1, 2018, to September 30, 2018 (Figure 1). The results of phases 1 and 3 have been published elsewhere [6,20]. This study aims to examine the results of phase 2 (ie, the results at the 18th month after the suspension of LEAN).

Figure 1. Study design [20]. LEAN: Lay health supporters, e-platform, award, and integration.



Procedures and Interventions

The development of our intervention *LEAN* was guided by the Health Belief Model [21-23], existing empirical evidence, and preliminary data from our pilot research [16]. On the basis of the Health Belief Model, patients' uptake of a health service depends on their evaluation of the risks of the health problem, the perceived barriers to and benefits of the action, and a *cue* or stimulus for action. Accordingly, the acronym of our intervention LEAN included four elements [6]: (1) lay health supporters (often a designated family member to watch patient medication, side effects, relapses, and urgent care), (2) e-platform (texting system for medication reminders, health education, and relapse monitoring), (3) award (token gifts to encourage behavioral improvement), and (4) integration of the texting with the existing health system to enable collaborative care. The e-platform was the core of the LEAN intervention. The educational and monitoring messages were intended to help the patients and families recognize the harm of schizophrenia and the benefits of taking medications. The SMS text messaging reminders serve as a *cue* to urge the patients to take action—essentially taking the medication.

Setting and Participants

The trial was conducted in a resource-poor setting: 9 rural townships of Liuyang municipality (population 356,900), Hunan

province, China. We included the patient participants if they (1) were living in the community rather than being institutionalized at the time of our recruitment, (2) were enrolled in the 686 Program, (3) had a confirmed diagnosis of schizophrenia according to International Classification of Diseases 10th Revision [24], (4) took oral psychotropic medications, and (5) resided in one of the nine rural townships of Liuyang municipality. People were excluded if they (1) were hospitalized for schizophrenia at the time of recruitment (our approach was community-based), (2) had missed 3 consecutive refills (in this case, they had de facto dropped out of the 686 Program), or (3) were physically incapable of using voice or SMS text messaging (hearing or vision impairment prevented the use of our intervention). Using simple random sampling, we drew a representative sample from the registry of the 686 Program, which included almost all known people with schizophrenia in Liuyang.

Outcomes

The details of the procedures to our outcome measurement and data collection were described in our published protocol and the earlier papers [6,16,20,25]. We provide a brief description in this section. We tracked all prespecified outcomes except the adherence measured by the Brief Adherence Rating Scale and the Drug Attitude Inventory-10 [16], which were not conducted as planned because of our errors in program implementation. The primary outcome was antipsychotic medication adherence as assessed by unannounced home-based pill counts (ie, the proportion of dosages taken in the past month). A total of 2 counts at a 30-day interval were performed to obtain the adherence using the following formula: (number of the first count − number of the second count + number of additional pills obtained − number of pills discarded) ÷ (number of pills prescribed). The count was considered *unannounced* as although the participants consented to the count, they were unaware of the specific timing of the count, which means that they were unaware of the first and second times of home visiting. The assessors (public health or medical students) were blinded to the group assignments and were rigorously trained to follow a pill count protocol that required inquiry on the number of purchased and discarded pills. The number of prescribed pills was abstracted from the 686 Program system. The pill count adherence was supplemented with medication refill adherence captured from the 686 Program system (number of refills required ÷ number of refills conducted over the past 6 months). We also measured two secondary outcomes: the patient's functioning assessed with the World Health Organization Disability Assessment Schedule 2.0 (WHODAS) [26,27] and symptoms assessed using the Clinical Global Impression–Schizophrenia (CGI-Sch) measure [28]. The student assessors administered the WHODAS in an interview with the patients. The 686 Program psychiatrists performed symptom evaluations using CGI-Sch. The 686 Program system routinely tracked a host of other outcomes, including rehospitalization due to schizophrenia, death for any reason, suicide, and wandering. We also collected data on violence in the past 6 months. In addition, we assessed at baseline a few other empirically suggested strong predictors of medication adherence, including medication side effects (assessed using the self-administered Glasgow Antipsychotic Side-Effect Scale [29]), smoking and alcohol abuse (ie, substance abuse), and family supervision on medicine [30].

Sample Size

The details of the sample size calculation were described by Xu et al [6,16]. Briefly, we calculated the sample size of 258 patient participants (129 per group) to ensure 85% power to detect the program effect of increasing the medication adherence from 0.72 to 0.85 (SD 0.33), which was considered minimally clinically acceptable difference after consultation with the 686 Program policy makers. The calculation assumed a 5% type I error and 10% attrition of participants.

Allocation and Blinding

A statistician who was not otherwise associated with the project assigned the participants equally to the intervention or control group using simple randomization. It was impossible to blind the program participants regarding their group assignment, but the outcome assessors and psychiatrists were blinded to the allocation status. The assessors were physically separated from the LEAN implementation team as well. Unmasking was reported immediately, and a makeup assessment was scheduled with separate assessors [16].

Statistic Methods

In this analysis, our primary objective was to analyze the sustained effect of LEAN at 18 months after the suspension of the LEAN intervention. Following our protocol, we used a generalized estimating equation model through the *gee( ) package* (gee; version 4.8) in R (RStudio; version 3.5.3) for statistical analyses. The generalized estimating equation model can provide a robust analysis of the abnormally distributed data of our primary outcome (adherence), which is the same as our phase1 analysis mode [6], so we can compare the difference between the 2 phases on the same model. Our primary outcome (adherence) analysis was adjusted for 7 covariates that were empirically suggested and prespecified baseline predictors of adherence in our protocol [16]. Multimedia Appendix 1 presents the primary and secondary analysis models and their equations.

We performed several sensitivity analyses to compare the results of the program's effects on adherence, functioning, and symptoms by fitting an unadjusted analysis without data imputation. We also fit a model with multiple imputation (MI) but without adjusted covariates and a model with both adjusted analysis with covariates and MI for the missing data. The details of MI are presented in Multimedia Appendix 2. We conducted the same analysis with covariate adjustment for the two prespecified subgroups: a baseline nonadherent group (missing any of the previous 6 refills) and the group with low baseline functioning (WHODAS: cutoff at ≤0.22). We performed an intention-to-treat analysis for all participants [31]. The fully conditional specification, an MI method, was used to impute missing data [32]. We calculated the program effect size as Cohen *d* [33] to enable a cross-study comparison.

Ethics and Dissemination

The study has obtained the institutional review board approval from the University of Washington (49464 G) and Central South

University (CTXY-150002-6). All patient participants and their lay health supporters provided written informed consent.

Patient and Public Involvement

Our study fully considered stakeholders' suggestions during the study design and implementation phases. The patients and family members directly informed our intervention. We piloted a system of asking the village doctors to observe the patients for medication taking directly. However, we later completely dropped this initial idea after consulting patients and doctors for their experiences in the pilot. They felt this was very intruding and added a reasonable burden to the village doctors. Taking the advice from the families, we revised our original plan and developed LEAN that relied on family members and mobile texting to help the patients. The stakeholders also helped us refine the frequency of texting, the role of lay health supporters, how the rewards would be delivered, and how the intervention would be integrated with the existing 686 Program. During the implementation phase, we also regularly captured the feedback from the stakeholders and used the information to further refine the program.

Protocol

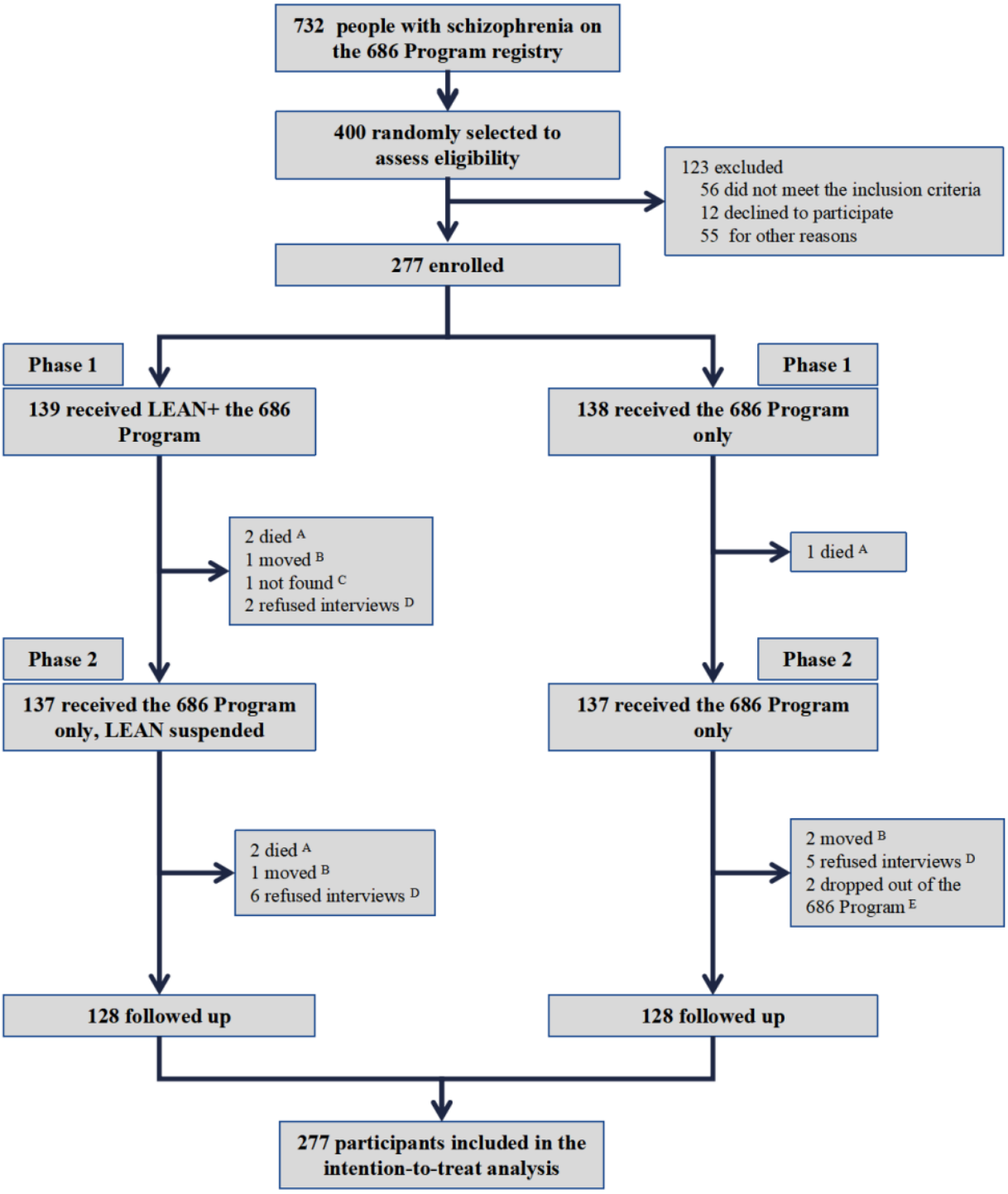
The study design, methods, and analysis plan have previously been published as a study protocol [16].

Results

Participants

We successfully recruited 277 patient participants (more than what we planned because of a higher level of interest): 139 randomized to the intervention group and 138 to the control group. Multimedia Appendix 3 summarizes the baseline characteristics between the intervention and control groups, suggesting no characteristic differences between the 2 groups. In phase 2, we collected the primary and secondary outcomes in 2 home visits conducted from November 24 to 29, 2017, and from December 24 to 29, 2017, respectively. The participants' flow is presented in Figure 2.

Figure 2. Participant flow [20]. (A) Died and thus no longer received the program; (B) moved but still received the program; (C) not found in the following interviews but still received the program; (D) refused interviews but still received the program; (E) dropped out of the program and thus no longer received the program. LEAN: Lay health supporters, e-platform, award, and integration.



Retention

Both the intervention and control groups had 128 (256/277, 92.4%) participants who completed phase 2 follow-up for the full range of information through home-based visits. However, we captured the data of 98.2% (272/277) of the participants for medication refill adherence when they came to the township health center to refill medicine. Among the 256 participants followed up by the home visit or visit the township center, 166 (64.8%) were successfully interviewed for pill count adherence, 247 (95.5%) for functioning (WHODAS), and 256 (100%) for symptoms (CGI-Sch). The analyses of the missing data pattern and the results of MIs are presented in Multimedia Appendix 2. The major results of this study are summarized in Table 1.

Table 1. Outcomes of different groups and periods in each phase.[a]

Measures and phase	Intervention group[b] (n=139)	Control group[b] (n=138)	Mean difference or relative risk[c] (95% CI)	*P* value
Primary outcome, mean (SD)				
Pill count adherence[d]				
1	0.61 (0.34)	0.48 (0.35)	0.11 (0.03 to 0.20)	.007
2[e]	0.62 (0.30)	0.58 (0.41)	0.05 (−0.06 to 0.16)	.41
Other measurements				
Refill record[f]				
1	0.83 (0.28)	0.76 (0.34)	0.08 (0.003 to 0.15)	.04
2	0.78 (0.34)	0.73 (0.37)	0.05 (−0.04 to 0.13)	.28
DAI-10[g,h]				
1	0.68 (0.20)	0.67 (0.22)	0.01 (−0.05 to 0.07)	.75
2	N/A[i]	N/A	N/A	N/A
BARS[j,k]				
1	0.71 (0.21)	0.68 (0.23)	0.03 (−0.03 to 0.08)	.38
2	N/A	N/A	N/A	N/A
Secondary outcome, mean (SD)				
WHODAS[l,m]				
1	0.12 (0.15)	0.14 (0.19)	−0.02 (−0.07 to 0.01)	.22
2[e]	0.16 (0.17)	0.20 (0.22)	−0.03 (−0.08 to 0.02)	.19
CGI-Sch[n] severity of illness[o]				
1	2.84 (1.37)	2.76 (1.24)	0.12 (−0.19 to 0.44)	.44
2[e]	2.50 (1.08)	2.58 (1.17)	−0.02 (−0.29 to 0.24)	.86
Negative				
1	2.61 (1.32)	2.85 (1.28)	−0.24 (−0.62 to 0.14)	.22
2	2.55 (1.04)	2.55 (1.13)	−0.01 (−0.27 to 0.26)	.97
Positive				
1	2.62 (1.32)	2.85 (1.24)	−0.24 (−0.63 to 0.14)	.22
2	2.51 (1.08)	2.60 (1.18)	−0.08 (−0.36 to 0.19)	.55
Depression				
1	2.25 (1.18)	2.06 (0.99)	0.19 (−0.13 to 0.51)	.25
2	2.60 (1.17)	2.65 (1.27)	−0.06 (−0.35 to 0.24)	.72
Cognition				
1	2.66 (1.31)	2.90 (1.25)	0.19 (−0.13 to 0.51)	.25
2	1.99 (0.92)	2.02 (0.98)	−0.02 (−0.25 to 0.21)	.84
CGI-Sch degree of change[p]				
1	3.09 (1.15)	3.02 (1.08)	0.04 (−0.23 to 0.32)	.76
2[e]	3.16 (1.13)	3.10 (1.20)	0.05 (−0.23 to 0.34)	.72
Other outcomes of the 686 Program[q], n (%)				
Rehospitalization due to schizophrenia				
1	9 (7.3)	25 (19.5)	0.37 (0.18 to 0.76)	.007

Measures and phase	Intervention group[b] (n=139)	Control group[b] (n=138)	Mean difference or relative risk[c] (95% CI)	*P* value
2	11 (8.4)	17 (12.9)	0.65 (0.32 to 1.33)	.24
Dead for any reason				
1	2 (1.4)	1 (0.8)	1.99 (0.18 to 21.65)	.57
2	2 (1.6)	0 (0)	N/A	N/A
Suicide				
1	0 (0)	0 (0)	N/A	N/A
2	1 (0.7)	0 (0)	N/A	N/A
Wander				
1	2 (1.4)	2 (1.5)	1.00 (0.96 to 1.06)	.98
2	1 (0.7)	1 (0.7)	1.01 (0.98 to 1.02)	.99
Hurting people or smashing objects				
1	2 (1.5)	6 (4.5)	0.37 (0.18 to 0.76)	.007
2	1 (0.7)	0 (0)	N/A	N/A
Making trouble				
1	0 (0)	0 (0)	N/A	N/A
2	1 (0.7)	0 (0)	N/A	N/A
Self-harm				
1	0 (0)	0 (0)	N/A	N/A
2	0 (0)	0 (0)	N/A	N/A

[a]Phase 1 results are cited from a published paper [6].

[b]The numbers in phases 1 and 2 represent the number of participants with complete information in the intervention and control groups, respectively, at the end of that phase.

[c]Other than other outcomes of the 686 Program, all outcomes are presented as mean differences. Meanwhile, pill count adherence, WHODAS, CGI-Sch severity of illness, and CGI-Sch degree of change are from the adjusted analysis that adjusted for baseline covariates and used data imputation for the missing data; all other outcomes were raw analysis results.

[d]The proportion of antipsychotic dosage taken over the past month assessed by unannounced home-based pillar counts (possible range 0-1).

[e]The results were from adjusted analysis with covariates and data imputation.

[f]Refill record adherence is (number of days medication obtained over the past 182 days) ÷ (182 days).

[g]DAI-10: Drug Attitude Inventory-10.

[h]Drug Attitude Inventory adherence was originally from −10 to +10 (higher score=more positive attitude toward medication), which was rescaled to 0 to 1.

[i]N/A: not applicable.

[j]BARS: Brief Adherence Rating Scale.

[k]BARS is the self-reported percentage of dosages administered over the past month.

[l]WHODAS: World Health Organization Disability Assessment Schedule 2.0.

[m]WHODAS: percentage of functioning loss (possible range 0-1).

[n]CGI-Sch: Clinical Global Impression–Schizophrenia.

[o]CGI-Sch severity of illness: higher scores indicate worse symptoms (possible range 1-7).

[p]CGI-Sch degree of change: higher scores indicate less change (possible range 1-7).

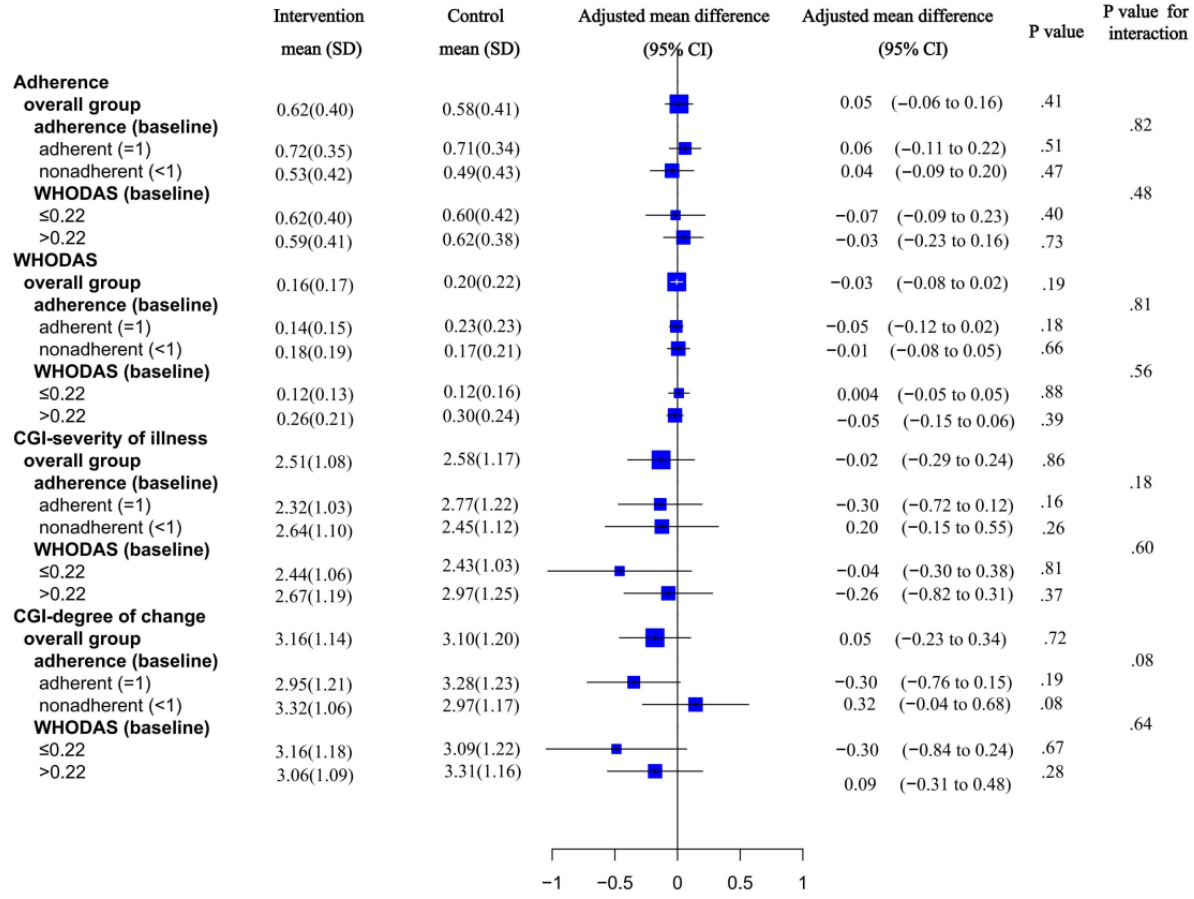
[q]These outcomes were tracked by the 686 Program administrative system on a routine basis. A small number of data points were missing.

Adherence Outcomes

In our study, the top 3 prescriptions of antipsychotic medications in phases 1 and 2 remained the same (Multimedia Appendix 4). In phase 1, the antipsychotic medication adherence improved after the intervention. However, in phase 2, there was no statistical significance of the difference of the antipsychotic medication adherence on pill counts between the intervention and control groups. In phase 1, the antipsychotic medication adherence based on pill counts increased from 0.48 (SD 0.35) in the control group to 0.61 (SD 0.34) in the intervention group in the adjusted analysis (adjusted mean difference 0.11, 95% CI 0.03-0.20; *P*=.007; Cohen *d* effect size=0.38; Table 1) [6]. However, the phase 2 results of this study indicated that, after the intervention was suspended, the difference in adherence between the 2 groups narrowed and was no longer significant. However, this narrowing effect for the antipsychotic medication adherence was mainly due to an increase in adherence in the

control group rather than a decrease in the intervention group (mean 0.62, SD 0.30 in the intervention group and mean 0.58, SD 0.45 in the control group; adjusted mean difference 0.05, 95% CI −0.06 to 0.16; *P*=.41; Cohen *d* effect size=0.11; Table 1). Subgroup analyses did not note significant improvement in phase 2 (Figure 3). We present the frequency of antipsychotic medication adherence and discontinuance reasons in phases 1 and 2. We also conducted a retrospective analysis of antipsychotic medication adherence in phase 1 for participants in phase 2. These results are presented in Multimedia Appendix 5, which shows that the antipsychotic medication adherence was statistically significantly different in the LEAN intervention group and control group in phase 1 (adjusted mean difference 0.12, 95% CI 0.01-0.21; *P*=.003).

Figure 3. Subgroup analyses. CGI: Clinical Global Impression; WHODAS: World Health Organization Disability Assessment Schedule 2.0.



Functioning

In phase 1, the mean scores of functioning were 0.12 (SD 0.15) and 0.14 (SD 0.19). There was no improvement in functioning. We still observed no improvement in functioning in phase 2, which has mean scores of 0.16 (SD 0.17) and 0.20 (SD 0.22; adjusted mean difference −0.03, 95% CI −0.08 to 0.02; *P*=.19; Cohen *d* effect size=0.20; Table 1). There was no significant improvement in functioning severity for the prespecified subgroups in phase 2 either (Figure 3).

Symptoms

In phase 1, there was no improvement in symptoms, with scores of CGI-Sch severity of illness of the 2 groups being 2.84 (SD 1.37) and 2.76 (SD 1.24). We did not note any improvement in CGI-Sch severity of illness, with scores between groups being 2.50 (SD 1.08) and 2.58 (SD 1.17; adjusted mean difference −0.02, 95% CI −0.29 to 0.24; *P*=.86; Cohen *d* effect size=0.07; Table 1), and CGI-Sch degree of change, with scores between groups being 3.16 (SD 1.13) and 3.10 (SD 1.20; adjusted mean difference 0.05, 95% CI −0.23 to 0.34; *P*=.72; Cohen *d* effect size=0.05; Table 1) in phase 2 either. There was no significant improvement in the symptoms for the prespecified subgroups in phase 2 either (Figure 3).

Other Outcomes

In phase 1, the incidence of rehospitalization due to schizophrenia decreased from 19.5% (25/128) under the control arm to 7.3% (9/124) under the intervention arm (relative risk 0.32, 95% CI 0.18-0.76; *P*=.007, number needed to treat 8.15, 95% CI 4.86-25.19; Table 1). However, the relative risk in rehospitalization in the original intervention group increased and was no longer significant, which was, however, mainly due to the reduced incidence of rehospitalization in the control arm. The incidence in the intervention group was 8.4% (11/131) and 12.9% (17/131) in the control group (relative risk 0.65, 95% CI 0.32-1.33; *P*=.24; number needed to treat 21.83, 95% CI 8.30-34.69; Table 1).

There was 1 patient reported for wandering, 1 for hurting people or smashing objects, 1 for making troubles, and 2 for death during the whole follow-up period, including 1 who committed suicide.

Sensitivity Analysis

Sensitivity analyses (Multimedia Appendix 6) showed that the results were not sensitive to the different analytical methods. The analyses of pill count adherence at various cutoff points also showed the same results—other cutoff points not affecting the results of the evaluation of the intervention after it was withdrawn ($P \geq .05$).

Discussion

Principal Findings

This study examined the sustaining effect of LEAN, a 2-arm randomized controlled trial of an SMS text messaging intervention, 18 months after its withdrawal. In phase 1 (the 6-month LEAN intervention), both the antipsychotic medication adherence based on pill counts and rehospitalization due to schizophrenia significantly improved (adherence from 0.48 to 0.61 and rehospitalization from 19.53% to 7.26%, control vs intervention) [6]. However, in phase 2 (18 months after the LEAN intervention was withdrawn), although adherence and rehospitalization in the intervention group almost remained at their phase 1 level, the control group participants' adherence improved, leading to the disappearance of the program advantage of LEAN plus the 686 Program versus the 686 Program alone. We did not observe any improvement in other outcomes or subgroups either.

Some studies have found that after the discontinuation of the texting intervention, the effect can last for a short or long time [17-19]. However, we should note several differences between our study and the other studies. We focused on people with schizophrenia who often have other cognitive impairment, whereas other studies targeted groups with better cognition, such as smokers, people with HIV, and adults with diabetes [17,18]. Meanwhile, compared with our work, many studies had shorter follow-up times, ranging from 8 weeks to 15 months after the intervention discontinuation [17-19]. The residual effect of LEAN can only be established by a significant mean difference in adherence between the initially assigned intervention group and the control group at the end point of the 18 months. The disappeared program advantage in our study was mainly due to the improvement in the control group rather than a diminishing adherence on the intervention group. We offer several speculations on the explanation of this seemingly puzzling result. First, some external conditions that affect medication adherence may have deferentially developed between the 2 groups. For instance, the intensity or quality of the 686 Program may have been applied differently to the 2 groups. However, our interviews with the families and program administrators did not support this. However, we cannot exclude other unknown factors. Second, there might have been a tortoise and hare effect [34] in that the discontinuation of texting reduced participants' motivations in the intervention group to keep improving, whereas the participants in the control group continued their gradual improvement at their original pace possible because of the management from the 686 Program. Third, there might be a ceiling effect that made it difficult for the adherence in the intervention group to continue improving. Fourth, the primary mechanism of LEAN may come from the texted *cues* to take medications rather than a changed belief among the participants on the benefits of taking medicines versus cost. Once this reminder is stopped, the program effect may disappear, partly because of the potential cognitive impairment in our participants. Finally, we cannot exclude an alternative explanation that the initial program advantage in phase 1 results from chance. Thus, the earlier texting did not help the patients and families establish the new behavior of taking medicine. At the same time, the visit of the data collection team for pill count might serve as a reminder for the families and patients to adhere to medications to improve the adherence in both groups. However, the effect might be more robust in the control group when the families found the adherence was rather poor.

After phase 2, we resumed the LEAN intervention for both the original intervention and control groups (phase 3). The published results of phase 3 suggest LEAN's effect on improving medication adherence over an extended implementation of LEAN (phase 3) [20]. No improvement in functioning has been observed throughout phases 1 [6], 2, and 3 [20]. In subgroup analyses, we did not find any functioning improvements in patients with different baseline adherence and functioning. We suspect the lack of effect was due to the ceiling effect. Because of the 686 Program that applied to both the intervention and control groups, people with schizophrenia in our study had much better functioning (WHODAS mean score 0.19) at baseline than other groups of people with schizophrenia of similar ethnicity and culture (WHODAS mean score ranging from 0.29 [35] to 0.64 [36]). People with schizophrenia in our study did not present a statistically significant improvement in symptoms between the intervention and control groups and in subgroups. However, we found a decreasing trend in CGI-Sch severity of illness scores and an increasing trend in CGI-Sch degree of change scores compared with themselves in phase 1. In phase 3, after an extended intervention, we found a statistically significant improvement in symptoms [20]. This indicates that sustained medication adherence may lead to improvement in symptoms.

Our study has several limitations, which may also point to future research directions. First, we should have conducted more systematic surveys and qualitative studies such as focus groups and interviews to understand participants' experiences in both the 686 Program and LEAN to better analyze the improvement in adherence in the control group 18 months after the LEAN suspension. Second, our failure in conducting the planned Drug Attitude Inventory (due to a scheduling error) at the end point of the 18-month period further complicated our efforts to understand the evolution of the participants' attitude toward the medications. Third, we did not systematically investigate the reasons for patients' adherence and nonadherence. Some validated scales such as the Antipsychotic Discontinuation Questionnaire may be considered [37-39]. Fourth, although our rigor validated pill count methods [25], measurement errors were possible (eg, patients intentionally discarding pills may

lead to inflated adherence measurement). However, this should not affect our program impact evaluation, as the behavior may occur for both groups and be balanced out because of our group randomization. Fifth, the low follow-up rate (166/277, 59.9%) for the primary adherence outcome was another limitation for our study. It resulted from the outcome-capturing method. We needed to interview the participants twice in 2 sequential months, and if there was 1 interview that failed, we could not count the difference of pills. If we informed the participants to wait for us at home, it would remind them to prepare for the pill counting, resulting in a false improvement in adherence. Finally, LEAN is a complex intervention because there were several program elements (ie, use of lay health supporters and program integration with the existing health system) besides the texted medication reminders. We cannot ascertain which component of the program delivered the largest effect. Neither can we test out the optimal dosage of texting in this trial. We strongly suggest future endeavors to consider trial designs proposed by implementation research, particularly those related to the Multiphase Optimization Strategy framework for the development of the complex intervention strategies and sequential multiple-assignment randomized trials [40] for the adaptive delivery of texting to people with schizophrenia [41]. Those new designs will refine the program elements of a complex intervention through systematically scheduled multiple phases and tailor the program intensity (dosage and contents of texting in this case) to different types of program participants. Quantitative comparative analysis may also be a valuable tool to understand the mechanism of the program [42]. In addition to these implementation research frameworks and methods, future studies may also use assessment frameworks for implementation outcomes such as Reach, Effectiveness, Adoption, Implementation and Maintenance to guide the implementation evaluation [43].

Conclusions

Our results provide new evidence and research directions related to maintaining adherence to medication after the intervention was terminated. Our study is inclined to suggest that continuous prompts for medicine may be necessary to support medication adherence for people with schizophrenia. However, we cannot definitively exclude other alternative interpretations.

Acknowledgments

The authors would like to thank all the participants: Liuyang Mental Health Center (Meng Dai and Li Chen) for providing patient assessment and policy support; mental health administrators from 9 township health centers (Zhong Huang, Huakun Li, Yanjiang Li, Xianyong Li, Jiaona Liu, Change Lu, Hao Luo, Xiao Pan, Huiren Shao, Shuyi Tang, Baojing Tian, Xiang Wang, and Qing Zeng) for coordinating and assisting in our fieldwork and home visits; the short message development and management group (Yongsheng Tong, Tianlin Ma, Junren Wang, Yun Luo, Jinan Long, and Jiali Ye); and the Lay health supporters, e-platform, award, and integration (LEAN) project volunteer group for field data collection and management. They were Chao Zhang, Min Yu, Qiong Tang, Yushi Mo, Anwen Yang, Bingwei Tang, Xin Jin, Yue Zeng, Weiyue Long, Huilan Tang, Wei Liu, Tingting Li, Xiaoyan Guo, Cheng Hu, Xing Liu, Nan Xia, Jingyan Fan, Jingjing Yao, Ziwei Ye, Hao Fang, Weina Zhang, Fan Yang, Wenqing Xue, Lu Tang, and Donglan Zhang.

The project received grant support from the China Medical Board (grant: 12-114, WG, principal investigator) and the National Institute on Aging of the National Institutes of Health (grant: P30AG047845, DRX, principal investigator). The funders had no role in the study design, data collection, decision to publish, or manuscript preparation.

Authors' Contributions

YC, DRX, WG, WH, SX, HH, JPH, JS, SG, JL, and YH conceptualized the study. XD, ZL, YC, and WH were responsible for data curation. YC, DRX, HH, WH, JPH, JS, SX, XD, ZL, BD, and WG conducted the formal analysis. DRX and WG were responsible for funding acquisition. YC, WH, and BD contributed to the investigation. DRX, HH, JPH, and WG contributed to the methodology. Project administration was the responsibility of DRX, YC, ML, and WG. The provision of resources was the responsibility of DRX, SX, and WG. The implementation of software-related work was handled by WH, XD, and ZL. DRX and WG supervised the work. JPH and HH contributed to the validation of data. ZL, WH, and XD prepared the data for visualization. YC and WG contributed equally to this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The analyses model and their equations in our study.
[DOCX File , 16 KB - mhealth_v10i4e33628_app1.docx]

Multimedia Appendix 2
Multiple imputation used in our statistical analysis.
[DOCX File , 235 KB - mhealth_v10i4e33628_app2.docx]

Multimedia Appendix 3
Baseline characteristics of participants.
[DOCX File , 15 KB - mhealth_v10i4e33628_app3.docx]

Multimedia Appendix 4
Prescribed antipsychotic medications in phases 1 and 2.
[DOCX File , 21 KB - mhealth_v10i4e33628_app4.docx]

Multimedia Appendix 5
Antipsychotic medication adherence and discontinuance reasons in phases 1 and 2.
[DOCX File , 153 KB - mhealth_v10i4e33628_app5.docx]

Multimedia Appendix 6
Sensitivity analysis for phase 2 outcomes.
[DOCX File , 26 KB - mhealth_v10i4e33628_app6.docx]

References

1. Iasevoli F, Giordano S, Balletta R, Latte G, Formato MV, Prinzivalli E, et al. Treatment resistant schizophrenia is associated with the worst community functioning among severely-ill highly-disabling psychiatric conditions and is the most relevant predictor of poorer achievements in functional milestones. Prog Neuropsychopharmacol Biol Psychiatry 2016;65:34-48. [doi: 10.1016/j.pnpbp.2015.08.010] [Medline: 26320028]
2. Phan SV. Medication adherence in patients with schizophrenia. Int J Psychiatry Med 2016;51(2):211-219. [doi: 10.1177/0091217416636601] [Medline: 27079779]
3. Bozzatello P, Bellino S, Rocca P. Predictive factors of treatment resistance in first episode of psychosis: a systematic review. Front Psychiatry 2019;10:67 [FREE Full text] [doi: 10.3389/fpsyt.2019.00067] [Medline: 30863323]
4. Leucht S, Cipriani A, Spineli L, Mavridis D, Orey D, Richter F, et al. Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. Lancet 2013;382(9896):951-962. [doi: 10.1016/S0140-6736(13)60733-3] [Medline: 23810019]
5. Lora A, Kohn R, Levav I, McBain R, Morris J, Saxena S. Service availability and utilization and treatment gap for schizophrenic disorders: a survey in 50 low- and middle-income countries. Bull World Health Organ 2012;90(1):47-54 [FREE Full text] [doi: 10.2471/BLT.11.089284] [Medline: 22271964]
6. Xu DR, Xiao S, He H, Caine ED, Gloyd S, Simoni J, et al. Lay health supporters aided by mobile text messaging to improve adherence, symptoms, and functioning among people with schizophrenia in a resource-poor community in rural China (LEAN): a randomized controlled trial. PLoS Med 2019;16(4):e1002785 [FREE Full text] [doi: 10.1371/journal.pmed.1002785] [Medline: 31013275]
7. Goff DC, Hill M, Freudenreich O. Strategies for improving treatment adherence in schizophrenia and schizoaffective disorder. J Clin Psychiatry 2010;71 Suppl 2:20-26. [doi: 10.4088/JCP.9096su1cc.04] [Medline: 21190649]
8. Good BJ, Good MD. Significance of the 686 Program for China and for global mental health. Shanghai Arch Psychiatry 2012;24(3):175-177 [FREE Full text] [doi: 10.3969/j.issn.1002-0829.2012.03.008] [Medline: 25324623]
9. Kang HM, Guan LL, Wang X, Wu XM. Economic condition, basic medical insurance and medication among patients with severe mental illness in 31 demonstration cities for 686 program implementation. Chin J Public Health 2015;31(8):997-1001. [doi: 10.11847/zgggws2015-31-08-06]
10. mhGAP: mental health gap action programme: scaling up care for mental, neurological and substance use disorders. World Health Organization. 2008. URL: https://www.who.int/mental_health/mhgap_final_english.pdf?ua=1 [accessed 2020-01-18]
11. Wu XM, Ma M, Wang X, Zhang WF, Guan L, Ma H, et al. Management and services for psychosis in the People's Republic of China in 2017. China J Psychiatry 2019;52(1):82-88. [doi: 10.3760/CMA.J.ISSN.1006-7884.2019.01.008]
12. Liang D, Mays VM, Hwang WC. Integrated mental health services in China: challenges and planning for the future. Health Policy Plan 2018;33(1):107-122 [FREE Full text] [doi: 10.1093/heapol/czx137] [Medline: 29040516]
13. Montes JM, Medina E, Gomez-Beneyto M, Maurino J. A short message service (SMS)-based strategy for enhancing adherence to antipsychotic medication in schizophrenia. Psychiatry Res 2012;200(2-3):89-95. [doi: 10.1016/j.psychres.2012.07.034] [Medline: 22901437]
14. Granholm E, Ben-Zeev D, Link PC, Bradshaw KR, Holden JL. Mobile Assessment and Treatment for Schizophrenia (MATS): a pilot trial of an interactive text-messaging intervention for medication adherence, socialization, and auditory hallucinations. Schizophr Bull 2012;38(3):414-425 [FREE Full text] [doi: 10.1093/schbul/sbr155] [Medline: 22080492]
15. Beebe L, Smith KD, Phillips C. A comparison of telephone and texting interventions for persons with schizophrenia spectrum disorders. Issues Ment Health Nurs 2014;35(5):323-329. [doi: 10.3109/01612840.2013.863412] [Medline: 24766166]

16. Xu DR, Gong W, Caine ED, Xiao S, Hughes JP, Ng M, et al. Lay health supporters aided by a mobile phone messaging system to improve care of villagers with schizophrenia in Liuyang, China: protocol for a randomised control trial. BMJ Open 2016;6(1):e010120 [FREE Full text] [doi: 10.1136/bmjopen-2015-010120] [Medline: 26792221]
17. Armanasco AA, Miller YD, Fjeldsoe BS, Marshall AL. Preventive health behavior change text message interventions: a meta-analysis. Am J Prev Med 2017;52(3):391-402. [doi: 10.1016/j.amepre.2016.10.042] [Medline: 28073656]
18. Dobson R, Whittaker R, Jiang Y, McNamara C, Shepherd M, Maddison R, et al. Long-term follow-up of a randomized controlled trial of a text-message diabetes self-management support programme, SMS4BG. Diabet Med 2020;37(2):311-318 [FREE Full text] [doi: 10.1111/dme.14182] [Medline: 31722130]
19. Menon V, Selvakumar N, Kattimani S, Andrade C. Therapeutic effects of mobile-based text message reminders for medication adherence in bipolar I disorder: are they maintained after intervention cessation? J Psychiatr Res 2018;104:163-168. [doi: 10.1016/j.jpsychires.2018.07.013] [Medline: 30081390]
20. Cai Y, Gong W, He H, Hughes JP, Simoni J, Xiao S, et al. Mobile texting and lay health supporters to improve schizophrenia care in a resource-poor community in rural China (LEAN trial): randomized controlled trial extended implementation. J Med Internet Res 2020;22(12):e22631 [FREE Full text] [doi: 10.2196/22631] [Medline: 33258788]
21. Rosenstock IM, Strecher VJ, Becker MH. Social learning theory and the Health Belief Model. Health Educ Q 1988;15(2):175-183. [doi: 10.1177/109019818801500203] [Medline: 3378902]
22. Finfgeld DL, Wongvatunyu S, Conn VS, Grando VT, Russell CL. Health belief model and reversal theory: a comparative analysis. J Adv Nurs 2003;43(3):288-297. [doi: 10.1046/j.1365-2648.2003.02712.x] [Medline: 12859788]
23. Becker MH, Maiman LA. Sociobehavioral determinants of compliance with health and medical care recommendations. Med Care 1975;13(1):10-24. [doi: 10.1097/00005650-197501000-00002] [Medline: 1089182]
24. The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. World Health Organization. Geneva, Switzerland: World Health Organization; 1992. URL: https://apps.who.int/iris/handle/10665/37958 [accessed 2021-11-30]
25. Xu DR, Gong W, Gloyd S, Caine ED, Simoni J, Hughes JP, et al. Measuring adherence to antipsychotic medications for schizophrenia: concordance and validity among a community sample in rural China. Schizophr Res 2018;201:307-314 [FREE Full text] [doi: 10.1016/j.schres.2018.05.014] [Medline: 29807806]
26. Sjonnesen K, Bulloch AG, Williams J, Lavorato D, B Patten S. Characterization of disability in Canadians with mental disorders using an abbreviated version of a DSM-5 emerging measure: the 12-item WHO disability assessment schedule (WHODAS) 2.0. Can J Psychiatry 2016;61(4):227-235 [FREE Full text] [doi: 10.1177/0706743716632514] [Medline: 27254415]
27. Guilera G, Gómez-Benito J, Pino O, Rojo JE, Cuesta MJ, Martínez-Arán A, et al. Utility of the World Health Organization disability assessment schedule II in schizophrenia. Schizophr Res 2012;138(2-3):240-247. [doi: 10.1016/j.schres.2012.03.031] [Medline: 22521724]
28. Haro JM, Kamath SA, Ochoa S, Novick D, Rele K, Fargas A, SOHO Study Group. The Clinical Global Impression-Schizophrenia scale: a simple instrument to measure the diversity of symptoms present in schizophrenia. Acta Psychiatr Scand Suppl 2003(416):16-23. [doi: 10.1034/j.1600-0447.107.s416.5.x] [Medline: 12755850]
29. Waddell L, Taylor M. A new self-rating scale for detecting atypical or second-generation antipsychotic side effects. J Psychopharmacol 2008;22(3):238-243. [doi: 10.1177/0269881107087976] [Medline: 18541624]
30. Fenton WS, Blyler CR, Heinssen RK. Determinants of medication compliance in schizophrenia: empirical and clinical findings. Schizophr Bull 1997;23(4):637-651. [doi: 10.1093/schbul/23.4.637] [Medline: 9366000]
31. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. BMJ 1999;319(7211):670-674 [FREE Full text] [doi: 10.1136/bmj.319.7211.670] [Medline: 10480822]
32. van Buuren S. Flexible imputation of missing data. Boca Raton, FL: CRC Press; 2012.
33. Cohen J. Statistical power analysis for the behavioral sciences. 2nd edition. New York, NY: Routledge; 1988.
34. Rangel-Malo RV, Molina-Lopez A, Jimenez-Tapia A, Lopez-Jimenez LA, Carriedo Garcia-Morato P, Gonzalez-Forteza CF. Changes after emergency assessment of suicidal patients: an unexpected outcome. Arch Suicide Res 2020:1-16. [doi: 10.1080/13811118.2020.1845888] [Medline: 33308106]
35. Chen R, Liou TH, Miao NF, Chang KH, Yen CF, Liao HF, et al. Using World Health Organization disability assessment schedule 2.0 in people with schizophrenia: a 4-year follow-up. Eur Arch Psychiatry Clin Neurosci 2020;270(3):301-310. [doi: 10.1007/s00406-019-01000-5] [Medline: 31011812]
36. Dan A, Kumar S, Avasthi A, Grover S. A comparative study on quality of life of patients of schizophrenia with and without depression. Psychiatry Res 2011;189(2):185-189. [doi: 10.1016/j.psychres.2011.02.017] [Medline: 21453977]
37. Matza LS, Phillips GA, Revicki DA, Ascher-Svanum H, Malley KG, Palsgrove AC, et al. Validation of a clinician questionnaire to assess reasons for antipsychotic discontinuation and continuation among patients with schizophrenia. Psychiatry Res 2012;200(2-3):835-842. [doi: 10.1016/j.psychres.2012.05.005] [Medline: 22841345]
38. Matza LS, Phillips GA, Revicki DA, Ascher-Svanum H, Malley KG, Palsgrove AC, et al. Validation of a patient interview for assessing reasons for antipsychotic discontinuation and continuation. Patient Prefer Adherence 2012;6:521-532 [FREE Full text] [doi: 10.2147/PPA.S25635] [Medline: 22879738]

39. Matza LS, Phillips GA, Revicki DA, Ascher-Svanum H, Kaiser D, Stauffer V, et al. Development of a clinician questionnaire and patient interview to assess reasons for antipsychotic discontinuation. Psychiatry Res 2011;189(3):463-468. [doi: 10.1016/j.psychres.2011.05.030] [Medline: 21684017]
40. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. Am J Prev Med 2007;32(5 Suppl):S112-S118 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.022] [Medline: 17466815]
41. Collins LM, Nahum-Shani I, Almirall D. Optimization of behavioral dynamic treatment regimens based on the sequential, multiple assignment, randomized trial (SMART). Clin Trials 2014;11(4):426-434 [FREE Full text] [doi: 10.1177/1740774514536795] [Medline: 24902922]
42. Hanckel B, Petticrew M, Thomas J, Green J. The use of Qualitative Comparative Analysis (QCA) to address causality in complex systems: a systematic review of research on public health interventions. BMC Public Health 2021;21(1):877 [FREE Full text] [doi: 10.1186/s12889-021-10926-2] [Medline: 33962595]
43. Yoshida Y, Patil SJ, Brownson RC, Boren SA, Kim M, Dobson R, et al. Using the RE-AIM framework to evaluate internal and external validity of mobile phone-based interventions in diabetes self-management education and support. J Am Med Inform Assoc 2020;27(6):946-956 [FREE Full text] [doi: 10.1093/jamia/ocaa041] [Medline: 32377676]

Abbreviations

LEAN: Lay health supporters, e-platform, award, and integration
CGI-Sch: Clinical Global Impression–Schizophrenia
MI: multiple imputation
WHODAS: World Health Organization Disability Assessment Schedule 2.0

Edited by L Buis; submitted 16.09.21; peer-reviewed by Y Tong, S Lalla-Edward, Y Yu, P Cook; comments to author 23.11.21; revised version received 12.01.22; accepted 02.02.22; published 19.04.22.

Please cite as:
Cai Y, Gong W, He W, He H, Hughes JP, Simoni J, Xiao S, Gloyd S, Lin M, Deng X, Liang Z, Dai B, Liao J, Hao Y, Xu DR
Residual Effect of Texting to Promote Medication Adherence for Villagers with Schizophrenia in China: 18-Month Follow-up Survey After the Randomized Controlled Trial Discontinuation
JMIR Mhealth Uhealth 2022;10(4):e33628
URL: https://mhealth.jmir.org/2022/4/e33628
doi:10.2196/33628
PMID:35438649

Original Paper

Japanese Version of the Mobile App Rating Scale (MARS): Development and Validation

Kazumichi Yamamoto[1,2*], MD, PhD; Masami Ito[1*], RN; Masatsugu Sakata[1], MA; Shiho Koizumi[3], BA; Mizuho Hashisako[4], PhD; Masaaki Sato[5], MD, PhD; Stoyan R Stoyanov[6], MRes; Toshi A Furukawa[1], MD, PhD

[1]Departments of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine, School of Public Health, Kyoto, Japan
[2]Research Unit, Institute for Airway Disease, Takarazuka, Japan
[3]Department of Health Informatics, Kyoto University Graduate School of Medicine, School of Public Health, Kyoto, Japan
[4]Department of Sociology, Rikkyo University, Tokyo, Japan
[5]Organ Transplantation Center, The University of Tokyo Hospital, Tokyo, Japan
[6]Institute of Health & Biomedical Innovation, School of Psychology and Counselling, Queensland University of Technology, Brisbane, Australia
[*]these authors contributed equally

Corresponding Author:
Kazumichi Yamamoto, MD, PhD
Departments of Health Promotion and Human Behavior
Kyoto University Graduate School of Medicine
School of Public Health
Yoshida-Konoe-Cho, Sakyo
Kyoto, 606-8501
Japan
Phone: 81 75 753 9492
Fax: 81 75 753 4641
Email: kazumichi_yamamoto@airwaystenosis.org

Abstract

Background: The number of mobile health (mHealth) apps continues to rise each year. Widespread use of the Mobile App Rating Scale (MARS) has allowed objective and multidimensional evaluation of the quality of these apps. However, no Japanese version of MARS has been made available to date.

Objective: The purposes of this study were (1) to develop a Japanese version of MARS and (2) to assess the translated version's reliability and validity in evaluating mHealth apps.

Methods: To develop the Japanese version of MARS, cross-cultural adaptation was used using a universalist approach. A total of 50 mental health apps were evaluated by 2 independent raters. Internal consistency and interrater reliability were then calculated. Convergent and divergent validity were assessed using multitrait scaling analysis and concurrent validity.

Results: After cross-cultural adaptation, all 23 items from the original MARS were included in the Japanese version. Following translation, back-translation, and review by the author of the original MARS, a Japanese version of MARS was finalized. Internal consistency was acceptable by all subscales of objective and subjective quality (Cronbach α=.78-.89). Interrater reliability was deemed acceptable, with the intraclass correlation coefficient (ICC) ranging from 0.61 to 0.79 for all subscales, except for "functionality," which had an ICC of 0.40. Convergent/divergent validity and concurrent validity were also considered acceptable. The rate of missing responses was high in several items in the "information" subscale.

Conclusions: A Japanese version of MARS was developed and shown to be reliable and valid to a degree that was comparable to the original MARS. This Japanese version of MARS can be used as a standard to evaluate the quality and credibility of mHealth apps.

(JMIR Mhealth Uhealth 2022;10(4):e33725) doi:10.2196/33725

KEYWORDS

mobile health apps; MHAs; mHealth; mobile application; mobile application rating scale; MARS; scale development; mental health; mobile health applications

Introduction

Smartphones are now an indispensable part of our lives. According to a 2021 global survey, more than 7.5 billion smartphones are in use around the world, and that number is only expected to increase [1]. With their growing popularity, they have come to have widespread applications in health care in many countries, including Japan [2]. The number of mobile health (mHealth) apps also continues to rise, especially since the beginning of the COVID-19 pandemic in early 2020 [3].

Although an increasing quantity of research showcases the efficacy of mHealth apps in many conditions, such as diabetes mellitus, asthma, and mental health [4-6], the overall evidence on their usefulness remains inconsistent [7]. This may reflect the lack of systematic research on the quality and efficacy of mHealth apps. [8,9].

To date, several medical societies [10,11] and researchers [12] have proposed ways to evaluate mHealth apps. Of these, the Mobile App Rating Scale (MARS) [13] is one of the most comprehensive, simple, and reliable. MARS is a 23-item scale, comprising 4 objective subscales and 1 subjective subscale (described in detail below). Validity and reliability are well supported for this scale [14], and an increasing number of studies use it to evaluate a wide range of mHealth apps [15-20].

The original version of MARS was developed in English, and several validated translations are available, including Italian, Spanish, German, French, and Arabic [21-25]. However, it has yet to be translated into any East Asian language, including Japanese, despite the recent increase in popularity of mHealth apps in this region. The development of standardized evaluation criteria shared among diverse cultures can contribute to the global public benefit of mHealth apps. Nevertheless, to date, no Japanese app evaluation scale exists.

The translation of scales involves not only a direct translation, but also adaptation of the questions to account for cultural differences, followed by appropriate measurements of reliability and validity [26]. The aims of this study were (1) to develop a Japanese version of MARS based on cross-cultural adaptation and (2) to assess the reliability and validity of this Japanese version by evaluating mHealth apps in the Japanese language.

Methods

Study Design

This study was conducted in two steps, following the methodology of previous translation and validation studies of the English MARS in other languages [21-25]: (1) cross-cultural adaptation with translation and back-translation and (2) a statistical evaluation of the reliability and validity of the translated scale.

MARS

The original MARS was developed by Stoyanov and colleagues [13] to establish a multidimensional measure able to classify and evaluate the objective and subjective quality of mHealth apps. The main part of this original version of MARS consisted of 23 items. The objective evaluation of mHealth app quality included 4 subscales: engagement (items 1-5), functionality (items 6-9), aesthetics (items 10-12), and information (items 13-19). The subjective quality subscale consisted of 4 items (items 20-23). Each MARS item is rated on a 5-point Likert scale (from 1 to 5: inadequate, poor, acceptable, good, and excellent), except for items 14 to 17 and item 19, which also have a "not applicable" option, for cases in which the item is not applicable to the evaluation. A mean score for each of the 4 objective subscales and an overall mean score of these 4 subscales are used. To determine subjective quality, individual scores for each item and a mean score for this subscale are rated separately. In addition to these 23 MARS items, sections to rate the classification, description, and perceived impact of the mHealth app can be adjusted according to the aims of the researcher. Both the original MARS [13] and several translated versions [21-25] have been assessed as providing high to satisfactory reliability and validity.

Cross-Cultural Adaptation and Translation Process

For the adaptation process, we were especially concerned about the cultural and linguistic differences between English and Japanese. Most of the existing translations of MARS are in European languages, which share some degree of cultural and linguistic similarities with English, but not with Japanese [27-29]. Therefore, we decided to adopt the "universalist" approach described by Herdman et al [30]. In this approach, 6 domains are considered for cross-cultural adaptation: item, conceptual, semantic, operational, measurement, and functional equivalence. Following these guidelines, each item and subcategory was assessed by a panel of 4 of the authors, comprising several specialties: a psychologist with a background in epidemiology (M Sakata), a registered nurse with a background in epidemiology (MI), a medical doctor and information technology developer (KY), and a sociologist specializing in questionnaire development (MH). All members are multilingual in Japanese, English, and other languages.

With the agreement of the panel, 3 translations were independently prepared by 3 panel members (M Sakata, MI, and KY). Following review and discussion of the differences between the 3 translations, a first draft of the Japanese translation was developed. This draft version was then back-translated into English, without referencing the original MARS scale or the original article, by a professional Japanese medical translator with a background in clinical epidemiology (SK). The back-translated version was proofread by a native English translator with a background in clinical pharmacy and clinical pharmacology. It was then reviewed and compared to the original by the developer of MARS (SS) and adjusted based on his feedback (Multimedia Appendix 1).

App Selection and Assessment

To better compare the results of this study with those of the original MARS, we tried to follow the original strategy for app selection and assessment. A systematic search was conducted on the Google Play Store and Apple App Store for mental health apps. The inclusion criteria were as follows: (1) the app was in Japanese, (2) the app was free, (3) the app was designed for adults, and (4) the app was developed by an entity based in Japan. The exclusion criteria were as follows: (1) the app

required the registration of personal information, (2) the app was unrelated to health, and (3) the app was developed for ongoing research by another academic entity. Because logic operators (AND, OR, and NOT) are not allowed in the Google Play Store or Apple App Store, the following keywords were used individually: “mindfulness,” “depression,” “wellbeing,” “well-being,” “mental health,” “anger,” “CBT,” “stress,” “distress,” and “anxiety.”

The sample size was calculated based on previous research [12,13,21]. A total of 41 apps were required to demonstrate interrater reliability within 0.15 of a sample observation of 0.80, with 87% assurance (based on 10,000 simulation runs) [31]. Ten apps were evaluated for the training stage, and to account for possible ineligible samples, a sample size of 60 apps was considered necessary for this study. If more than 60 apps were eligible after the systematic search, 60 apps were randomly selected using a random sequence. If an app turned out to be ineligible for evaluation, it was eliminated and another app randomly selected from among the remaining eligible apps.

After watching a training video provided by the author of the original MARS (SS), 10 apps were rated independently by 3 raters (MI, KY, M Sakata) as a training exercise. Then, disagreements were discussed until a consensus was reached to ensure consistent interpretation of all MARS terminology and item logic. Two raters independently assessed the remaining 50 apps in the final analysis.

Statistical Analysis

Descriptive Statistics

The distribution of summary scores (for the total and subscale scores for objective quality) was visually inspected and evaluated for a normal distribution using skewness and the Shapiro-Wilk test. Skewness was judged significant if the estimate was more than plus or minus 1.0. Normally distributed data were expressed as the mean (SD). Floor or ceiling effects were judged to be present if more than 15% of the apps were rated as the lowest or highest scores, respectively.

Reliability

The internal consistency of the total and subscale scores for objective quality was assessed using Cronbach α. Internal consistency was deemed acceptable at $\alpha > .6$ [32]. The interrater reliability was assessed using the intraclass correlation coefficient (ICC) using 2-way mixed effects and an averaged-measurements model with absolute agreement [13,21,22]. ICC was judged acceptable at >0.5 [33].

Validity

For construct validity, item-subscale correlations were investigated using multitrait scaling analysis [34]. The convergent validity was deemed satisfactory if the item achieved at least a correlation of 0.4 with its item-own subscale. For discriminant validity, the correlation coefficients of each item with an item-own subscale were compared with those with other subscales. The discriminant validity was considered satisfactory if more than 80% of correlation coefficients in the item-own subscale were higher than those with other subscales [22]. We expressed these estimates as the success rate—the number of items that fulfilled the above-mentioned conditions, divided by the total number of items within the subscale. This success rate was only calculated for the 4 objective quality subscales, because subjective quality is rated independently from objective quality in MARS.

To determine concurrent validity, the lack of an external “gold standard” rating scale led us to compare the correlation between the mean scores from 4 subscales of objective quality against the star rating and subjective quality total mean score using the Pearson *r* coefficient with 95% CI. The correlation between the mean total score of objective quality and mean star ratings in the app stores was also determined as in the original MARS [13].

Statistical Software

R (version 4.0.5; R Foundation for Statistical Computing) was used for all analyses.

Results

Cross-Cultural Adaptation and Translation Process

The 4 specialists held a joint discussion to conduct a conceptual analysis of the Japanese translation. All subscales and items were evaluated for conceptual equivalence between English and Japanese. The panel agreed to include all items in all of the subscales in the translation.

No major discrepancies were found among the 3 independently developed translations. All differences in expression were resolved through discussion. However, we encountered issues when translating several words that had no Japanese equivalent. For example, for the word “engagement,” it seemed that no Japanese word could express this concept. In such cases, we translated the word into terms as close as possible to the original concept together with the phonetic rendition in *katakana*, a Japanese syllabary used to express foreign words based on their pronunciation.

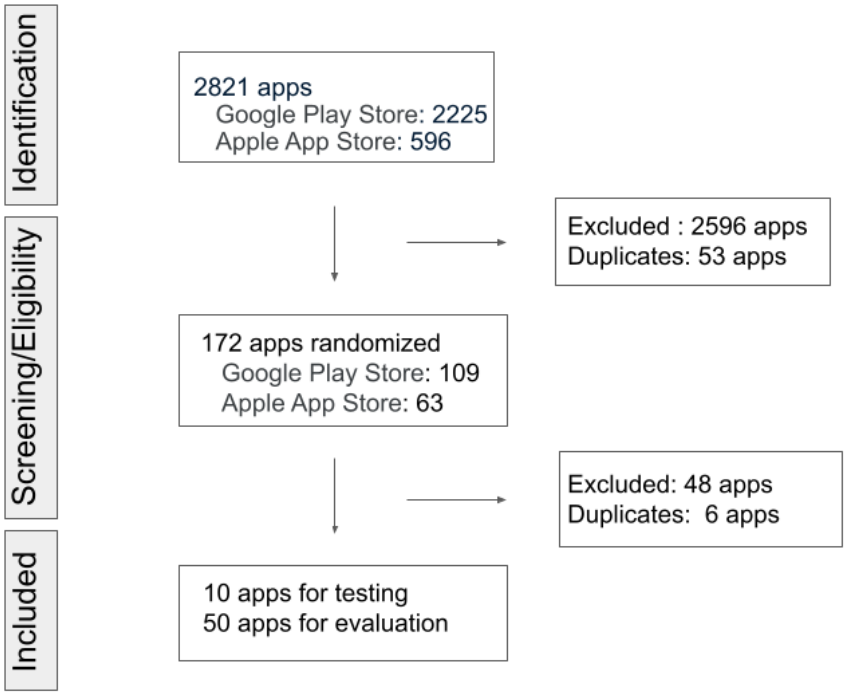
After creating the initial Japanese version of MARS, a back-translation was sent to the author of the original MARS without modifications. In general, the back-translated version was deemed equivalent to the original MARS. Several comments were provided to clarify word meanings. All comments from the original MARS author were reviewed and integrated, where relevant, by the 4 researchers who developed the Japanese MARS together with the translator who provided the back-translation (SK). Because the back-translation was considered appropriate and no major changes were made, no second back-translation was created after discussion with the author of the original MARS (SS).

App Selection and Test Phase

A search of the Apple App Store and Google Play Store was conducted on June 4 and June 11, 2021. A total of 2821 apps (Apple App Store: n=596; Google Play Store: n=2225) were retrieved. All the apps were screened for adherence to the inclusion and exclusion criteria based on the information page for the app. Of 225 candidate apps, 53 were duplicates, and the remaining 172 apps were the final candidates for random sampling. A computer-generated random sequence was assigned,

and the first 60 apps were selected for testing and evaluation (Figure 1). Fifty-four apps were excluded from the list during this rating phase based on the inclusion and exclusion criteria.

Figure 1. Flow diagram showing the process of identifying apps for pilot use of the Mobile App Rating Scale (MARS).



Reliability and Validity Analysis

Among the 50 apps analyzed, 36 (72%) were from the Google Play Store, and 14 (28%) were from the Apple App Store. A response of "not applicable (N/A)" was allowed for items 14 to 17 and 19 when there were no concrete goals (item 14), no information (items 15-17), or no search results in Google Scholar (item 19). More than 50% of the values for these items were therefore missing (73%, 60%, 61%, 58%, and 91% for items 14 to 17 and 19, respectively). It was decided to treat these values as missing in most of the analyses, except for the item-subscale correlation analysis, where a value of zero was assigned as "not applicable."

Table 1 shows the descriptive analysis results. No skewness was apparent in subscale score distributions. The Shapiro-Wilk test revealed a lack of fit to a normal distribution in several subscales. However, after visual inspection of the distributions, the mean (SD) was finally determined for all subscales. No ceiling or floor effects were detected.

Table 2 shows the results of the reliability analysis. Cronbach α was deemed acceptable in all objective and subjective quality subscales, with a range of α=.78 to .89. ICC results were considered acceptable for all subscales of objective quality and subjective quality, falling within the range of 0.61 to 0.79, except for the "functionality" subscale, which had an ICC of 0.40 (95% CI 0.20-0.54).

As shown in Table 3, the results of convergent and divergent validity were analyzed using multitrait scaling analysis. Item 19 was eliminated from the analysis because more than 90% of responses were "not applicable." As for convergent validity, most items were deemed acceptable with a correlation of >0.4, and the success rate was satisfactory, except for the subscale "information" (50%). For divergent validity, most items were satisfactory, with more than an 80% success rate, except for the subscale "information" (67%). Figure 2 shows a visual image of item-subscale relationships in subscales of objective quality.

Table 4 shows the concurrent validity based on the Pearson correlation coefficient between the total score (ie, the combined scores for objective and subjective quality) vs the MARS star rating (item 23) and the star rating on the app stores (ie, Google Play Store and Apple App Store). A statistically significant correlation was found between the total score and the MARS star rating at >0.8 with a relatively narrow 95% CI. However, this correlation was not observed in the correlation between the total score and the app store star rating (0.17-0.3), which had a wider 95% CI.

Table 1. Descriptive statistics.

Scale	Skewness	Shapiro-Wilk (*P*)	Ceiling effect (%)	Floor effect (%)	Mean (SD)
Objective quality					
Engagement	0.25	0.98 (.16)	1	2	2.64 (0.74)
Functionality	–0.96	0.93 (<.001)	2	2	3.67 (0.82)
Aesthetics	0.21	0.96 (.002)	4	3	3.13 (0.83)
Information	–0.29	0.97 (.06)	1	2	2.98 (0.69)
Total Score	–0.16	0.99 (.32)	1	1	2.90 (0.63)
Subjective quality	0.53	0.93 (<.001)	1	14	2.20 (0.94)

Table 2. Internal consistency and interrater reliability.

Scale	Cronbach α	Intraclass correlation coefficient (95% CI)
Objective quality		
Engagement	.78	0.69 (0.57-0.77)
Functionality	.83	0.40 (0.20-0.54)
Aesthetics	.89	0.61 (0.4-0.72)
Information	.82	0.79 (0.23-0.75)
Total Score	.81	0.70 (0.65-0.74)
Subjective quality	.88	0.75 (0.67–0.81)

Table 3. Construct validity measured with multitrait scaling analysis.

Subscale and item	Corrected item-subscale correlation	Success rate[a]	
		Convergent validity	Divergent validity
Engagement		4/5	4/5
Item 1	0.35		
Item 2	0.61		
Item 3	0.65		
Item 4	0.53		
Item 5	0.62		
Functionality		4/4	4/4
Item 6	0.59		
Item 7	0.55		
Item 8	0.81		
Item 9	0.73		
Aesthetics		3/3	3/3
Item 10	0.68		
Item 11	0.84		
Item 12	0.83		
Information		3/6	4/6
Item 13	0.24		
Item 14	0.33		
Item 15	0.74		
Item 16	0.75		
Item 17	0.39		
Item 18	0.49		
Item 19[b]	—[c]	—	—
Subjective quality[d]			
Item 20	0.83	—	—
Item 21	0.84	—	—
Item 22	0.55	—	—
Item 23	0.78	—	—

[a]Success rate was defined as the rate of prespecified acceptable items among all items in each subscale.

[b]Item 19 was eliminated from the analysis because of missing values.

[c]Not applicable.

[d]Success rate was not calculated for subjective quality.

Figure 2. Box plots of subscale correlations with item-own and other subscales. The mean correlation of each subscale is higher than the correlation with other subscales.

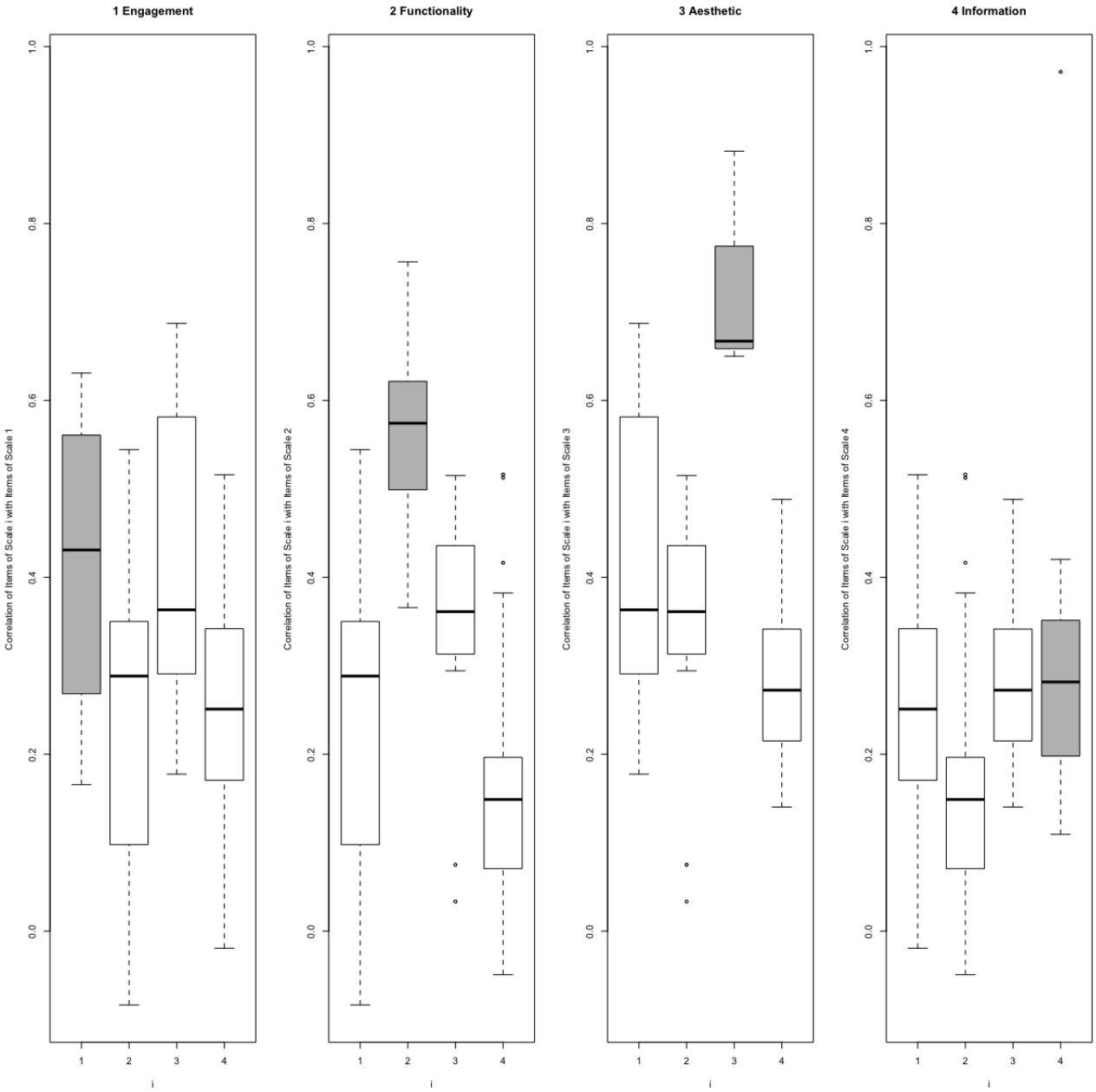


Table 4. Concurrent validity of total score measured with the Pearson correlation coefficient.

Scale	Pearson *r*	95% CI	*P* value
Total score vs subjective quality	0.85	0.79-0.90	<.001
Total score vs star rating (item 23)	0.84	0.77-0.89	<.001
Total score vs star rating (app stores)	0.24	0.03-0.42	.02

Discussion

Main Study

To our knowledge, this is the first time a cross-cultural approach has been used in the development and validation of a Japanese version of the MARS. This study also includes the involvement of one of the authors of the original MARS. It provides a statistical evaluation of the reliability and validity of the Japanese version in assessing 50 apps in Japanese.

We adopted the universalist approach [30] following practices from previous studies on translations and cross-cultural validation of MARS in other languages. There is controversy about whether it is preferable to adopt a universalist or country-specific approach to patient-reported outcome measures. However, the consortium that qualifies patient-reported

outcomes for use in clinical trials in the United States prefers a universalist approach to minimize the variability of language-related logistical complexity [35]. The universalist approach has substantial advantages in achieving conceptual equivalence in cross-cultural translation. Following this approach, we formed a panel with members from a wide variety of disciplines, not only limited to medicine or psychology, but also including an information technology developer, a professional translator with a background in epidemiology, and a sociologist specializing in the development of questionnaires. After discussion within the panel to account for linguistic differences between Japanese and the original English version, we finally decided to include all items with minor modifications for the Japanese version, based on conceptual, semantic, operational, measurement, and functional equivalence. For a cultural adaptation, we believe that it is most practical and helpful to involve specialists from a broad range of backgrounds.

During app selection, 172 apps were found eligible for evaluation, of which 60 were selected. Surprisingly, more than 90% of these apps lacked any scientific evidence supporting them; we found neither research nor supporting articles on their efficacy. Takashina et al [36] evaluated 47 apps for depression that had been developed in Japanese and concluded that very few apps were evidence-based and secure. This situation is quite problematic, because inaccurate or misleading apps could potentially impair the health of users or lead to incorrect decision-making [22]. For this reason, the current study offers a step in the right direction by translating a well-established quality evaluation scale into Japanese.

The analysis of reliability and validity suggests that our results are comparable with the original MARS and other translated versions [13,21-24]. The internal consistency and Cronbach α of all subscales and total scores were satisfactory according to the internationally established quality criteria [37]. This high internal consistency was also observed in the original MARS study and in previous translation studies. Conversely, our study showed slight variability in interrater reliability, with an ICC in the range of 0.40 to 0.79. This finding was also observed in the original MARS, which had a range of 0.50 to 0.83. We evaluated 10 apps after watching a training video provided by the author of the original MARS; 2 raters then discussed the evaluation. Disagreements were discussed until a consensus was reached. We still found low ICC for the "functionality" subscale, however. In that sense, we consider that a test phase and use of a training video are particularly important in assuring mHealth apps are rated correctly.

As for construct validity, we used a multitrait scaling analysis with item-subscale correlation rather than a factor analysis, because this method has been successfully applied in all previous studies. Our results were satisfactory in terms of convergent/divergent validity and fulfilled the prespecified success level, except for the "information" subscale, in which "not applicable" was the choice for most items. This was assigned a value of zero instead of being reported as a missing value. As in the original MARS study, the Japanese version also accepts "not applicable" as a response to items 14 to 17 and 19. During the evaluation, we frequently encountered apps where no clear goal was stated and no information on the source or detailed explanations were provided. In these cases, "not applicable" was chosen rather than one of the choices of the Likert scale. MARS itself takes such situations into account and uses the mean of the subscale total score. However, this is a problem for validation because the proportion of "not applicable" answers exceeded 50% for items 14 to 17 and 19. As a way of resolving this, we assigned zero as the numerical score for "not applicable" rather than treating these as missing values in items 14 to 17, thus allowing a comparison of the proximity of the scores between the raters. Item 19 was eliminated from the analysis, as it was in other MARS translation studies, because mHealth apps mostly lack evidence-based evaluation research, which the item aims to measure. We believe this should be clarified in a future updated version once a better practice for mHealth evaluation is widely implemented.

When measuring concurrent validity, the MARS objective quality total score was significantly and closely correlated with the subjective quality total score and star rating (item 23), with Pearson r>0.8. However, it was fairly well correlated with the star rating on the app stores. This finding has also been seen in previous studies [13,22]. As Stoyanov et al [13] reported, it is possible that the MARS subjective quality rating may be influenced by the completion of the MARS objective quality rating, and the results should be evaluated with caution. However, the lack of reliability of the star ratings on app stores has also been reported [38], and in this sense, MARS subjective score or star ratings could be a more reliable indicator of the ratings of mHealth apps.

Limitations

This study has several limitations. First, this was a validation study that tested only mental health apps. This was to maintain comparability with the original MARS study, which also studied only mental health mHealth apps. However, other translated MARS studies have used apps on other topics, such as physical activity [21] and primary prevention [22]. Accumulating evidence in recent publications shows that MARS is being used to evaluate mHealth apps in a wide variety of areas [15-20]. Thus, availability of a Japanese MARS will facilitate further research on app validation. Secondly, as mentioned above, the "information" subscale could not be adequately validated in this study. Neither the original MARS study nor other translation studies have had missing values, except for item 19, which estimates the degree of the evidence base of an app. However, several items do allow the "not applicable" choice and there are no clearly defined guidelines on the appropriate use of this rating option in the original MARS version. For this reason, it may be prudent to specify standards on choosing this option in future updated versions.

Future Research

Based on the results of the present study, we would like to propose several topics for future research. First, as made apparent in this study, validation requires further research. In almost all previous studies, item 19 (ie, the evidence base) was excluded from the analysis because of missing data. MARS was created to take missing values into account and uses mean scores instead of sum scores. This, however, makes it complicated to

estimate the validity of individual items; more research needs to be performed to validate the scale. Second, a more detailed validation of the Japanese version of MARS is also required, especially regarding app classification and perceived impact. In the present study, we only validated the main MARS components. Lastly, the goal of mHealth apps should be to improve health outcomes. As the present and previous studies show, few mHealth apps have been evaluated and assessed in medical studies. This means that most mHealth apps lack any evidence on health outcomes. In this sense, health outcome improvements through the use of mHealth apps need to be evaluated using standardized measures, such as randomized controlled trials. MARS can be used in conjunction with such studies to help determine the link between app quality and efficacy.

Conclusion

A Japanese version of MARS was developed and shown to be as reliable and valid as the original MARS. The Japanese version of MARS can be used as a standard in evaluating the quality and credibility of mHealth apps. Further research is required for additional validation and for exploring the application of the scale in a range of research contexts.

Acknowledgments

Author KY received support from Grants-in-Aid for Scientific Research of the Japan Society for Promotion of Science (20K18881). We thank Dr Sako Ikegami for proofreading the back-translation and the final manuscript.

Conflicts of Interest

TAF reports grants and personal fees from Mitsubishi-Tanabe, personal fees from MSD, personal fees from Sony, and grants and personal fees from Shionogi, all of which were outside the submitted work. In addition, TAF has a pending patent (2020-548587) concerning smartphone cognitive behavioral therapy apps and intellectual properties for the Kokoro app, which is licensed to Mitsubishi-Tanabe. SK received a back-translation fee from the Institute for Airway Disease. M Sakata reports personal fees from Sony. No other authors have any conflicts of interest.

Multimedia Appendix 1
Japanese version of the Mobile Application Rating Scale (MARS-Japanese).
[PDF File (Adobe PDF File), 206 KB - mhealth_v10i4e33725_app1.pdf]

References

1. Smartphone users 2021. Statistia. URL: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/ [accessed 2021-09-07]
2. Ministry of Internal Affairs and Communications. Information and Communication White Paper 2019. URL: https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r01/html/nd232110.html [accessed 2021-09-07]
3. COVID-19 growth in medical app downloads by country 2020. Statistia. URL: https://www.statista.com/statistics/1181413/medical-app-downloads-growth-during-covid-pandemic-by-country/ [accessed 2021-09-07]
4. Eberle C, Löhnert M, Stichling S. Effectiveness of Disease-Specific mHealth Apps in Patients With Diabetes Mellitus: Scoping Review. JMIR Mhealth Uhealth 2021 Feb 15;9(2):e23477 [FREE Full text] [doi: 10.2196/23477] [Medline: 33587045]
5. Snoswell CL, Rahja M, Lalor AF. A Systematic Review and Meta-Analysis of Change in Health-Related Quality of Life for Interactive Telehealth Interventions for Patients With Asthma. Value Health 2021 Feb;24(2):291-302. [doi: 10.1016/j.jval.2020.09.006] [Medline: 33518036]
6. Lecomte T, Potvin S, Corbière M, Guay S, Samson C, Cloutier B, et al. Mobile Apps for Mental Health Issues: Meta-Review of Meta-Analyses. JMIR Mhealth Uhealth 2020 May 29;8(5):e17458 [FREE Full text] [doi: 10.2196/17458] [Medline: 32348289]
7. Iribarren SJ, Akande TO, Kamp KJ, Barry D, Kader YG, Suelzer E. Effectiveness of Mobile Apps to Promote Health and Manage Disease: Systematic Review and Meta-analysis of Randomized Controlled Trials. JMIR Mhealth Uhealth 2021 Jan 11;9(1):e21563. [doi: 10.2196/21563]
8. The Lancet. Does mobile health matter? The Lancet 2017 Nov;390(10109):2216. [doi: 10.1016/S0140-6736(17)32899-4]
9. Jake-Schoffman DE, Silfee VJ, Waring ME, Boudreaux ED, Sadasivam RS, Mullen SP, et al. Methods for Evaluating the Content, Usability, and Efficacy of Commercial Mobile Health Apps. JMIR Mhealth Uhealth 2017 Dec 18;5(12):e190 [FREE Full text] [doi: 10.2196/mhealth.8758] [Medline: 29254914]
10. The App Evaluation Model. American Psychiatric Association. URL: https://www.psychiatry.org/psychiatrists/practice/mental-health-apps/app-evaluation-model [accessed 2020-07-15]
11. Mental Health Apps: How to Make an Informed Choice. Mental Health Commission of Canada. 2018. URL: https://www.mentalhealthcommission.ca/sites/default/files/2018-01/eMH_app_eng.pdf [accessed 2020-07-15]

12. Nouri R, R Niakan Kalhori S, Ghazisaeedi M, Marchand G, Yasini M. Criteria for assessing the quality of mHealth apps: a systematic review. J Am Med Inform Assoc 2018 Aug 01;25(8):1089-1098. [doi: 10.1093/jamia/ocy050] [Medline: 29788283]
13. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
14. Terhorst Y, Philippi P, Sander LB, Schultchen D, Paganini S, Bardus M, et al. Validation of the Mobile Application Rating Scale (MARS). PLoS One 2020;15(11):e0241480 [FREE Full text] [doi: 10.1371/journal.pone.0241480] [Medline: 33137123]
15. Davalbhakta S, Advani S, Kumar S, Agarwal V, Bhoyar S, Fedirko E, et al. A Systematic Review of Smartphone Applications Available for Corona Virus Disease 2019 (COVID19) and the Assessment of their Quality Using the Mobile Application Rating Scale (MARS). J Med Syst 2020 Aug 10;44(9):164 [FREE Full text] [doi: 10.1007/s10916-020-01633-3] [Medline: 32779002]
16. Kim BY, Sharafoddini A, Tran N, Wen EY, Lee J. Consumer Mobile Apps for Potential Drug-Drug Interaction Check: Systematic Review and Content Analysis Using the Mobile App Rating Scale (MARS). JMIR Mhealth Uhealth 2018 Mar 28;6(3):e74 [FREE Full text] [doi: 10.2196/mhealth.8613] [Medline: 29592848]
17. Romero RL, Kates F, Hart M, Ojeda A, Meirom I, Hardy S. Quality of Deaf and Hard-of-Hearing Mobile Apps: Evaluation Using the Mobile App Rating Scale (MARS) With Additional Criteria From a Content Expert. JMIR Mhealth Uhealth 2019 Oct 30;7(10):e14198 [FREE Full text] [doi: 10.2196/14198] [Medline: 31670695]
18. Mandracchia F, Llauradó E, Tarro L, Valls RM, Solà R. Mobile Phone Apps for Food Allergies or Intolerances in App Stores: Systematic Search and Quality Assessment Using the Mobile App Rating Scale (MARS). JMIR Mhealth Uhealth 2020 Sep 16;8(9):e18339 [FREE Full text] [doi: 10.2196/18339] [Medline: 32936078]
19. Amor-García, Collado-Borrell R, Escudero-Vilaplana V, Melgarejo-Ortuño A, Herranz-Alonso A, Arranz Arija, et al. Assessing Apps for Patients with Genitourinary Tumors Using the Mobile Application Rating Scale (MARS): Systematic Search in App Stores and Content Analysis. JMIR Mhealth Uhealth 2020 Jul 23;8(7):e17609 [FREE Full text] [doi: 10.2196/17609] [Medline: 32706737]
20. Escriche-Escuder A, De-Torres I, Roldán-Jiménez C, Martín-Martín J, Muro-Culebras A, González-Sánchez M, et al. Assessment of the Quality of Mobile Applications (Apps) for Management of Low Back Pain Using the Mobile App Rating Scale (MARS). Int J Environ Res Public Health 2020 Dec 09;17(24):9209 [FREE Full text] [doi: 10.3390/ijerph17249209] [Medline: 33317134]
21. Martin Payo R, Fernandez Álvarez MM, Blanco Díaz M, Cuesta Izquierdo M, Stoyanov SR, Llaneza Suárez E. Spanish adaptation and validation of the Mobile Application Rating Scale questionnaire. Int J Med Inform 2019 Sep;129:95-99. [doi: 10.1016/j.ijmedinf.2019.06.005] [Medline: 31445295]
22. Domnich A, Arata L, Amicizia D, Signori A, Patrick B, Stoyanov S, et al. Development and validation of the Italian version of the Mobile Application Rating Scale and its generalisability to apps targeting primary prevention. BMC Med Inform Decis Mak 2016;16:83 [FREE Full text] [doi: 10.1186/s12911-016-0323-2] [Medline: 27387434]
23. Messner E, Terhorst Y, Barke A, Baumeister H, Stoyanov S, Hides L, et al. The German Version of the Mobile App Rating Scale (MARS-G): Development and Validation Study. JMIR Mhealth Uhealth 2020 Mar 27;8(3):e14479 [FREE Full text] [doi: 10.2196/14479] [Medline: 32217504]
24. Saliasi I, Martinon P, Darlington E, Smentek C, Tardivo D, Bourgeois D, et al. Promoting Health via mHealth Applications Using a French Version of the Mobile App Rating Scale: Adaptation and Validation Study. JMIR Mhealth Uhealth 2021 Aug 31;9(8):e30480 [FREE Full text] [doi: 10.2196/30480] [Medline: 34463623]
25. Bardus M, Awada N, Ghandour LA, Fares E, Gherbal T, Al-Zanati T, et al. The Arabic Version of the Mobile App Rating Scale: Development and Validation Study. JMIR Mhealth Uhealth 2020 Mar 03;8(3):e16956 [FREE Full text] [doi: 10.2196/16956] [Medline: 32130183]
26. Sousa VD, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. J Eval Clin Pract 2011 Apr;17(2):268-274. [doi: 10.1111/j.1365-2753.2010.01434.x] [Medline: 20874835]
27. Gray RD, Atkinson QD. Language-tree divergence times support the Anatolian theory of Indo-European origin. Nature 2003 Nov 27;426(6965):435-439. [doi: 10.1038/nature02029] [Medline: 14647380]
28. Robbeets M, Bouckaert B. Bayesian phylo linguistics reveals the internal structure of the Transeurasian family. Journal of Language Evolution, Jul;? 2018;3(2):62. [doi: 10.1093/jole/lzy007]
29. Coupé C, Oh YM, Dediu D, Pellegrino F. Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche. Sci Adv 2019 Sep;5(9):eaaw2594 [FREE Full text] [doi: 10.1126/sciadv.aaw2594] [Medline: 32047854]
30. Herdman M, Fox-Rushby J, Badia X. A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. Qual Life Res 1998 May;7(4):323-335. [doi: 10.1023/a:1024985930536] [Medline: 9610216]
31. Zou GY. Sample size formulas for estimating intraclass correlation coefficients with precision and assurance. Stat Med 2012 Dec 20;31(29):3972-3981. [doi: 10.1002/sim.5466] [Medline: 22764084]

32. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951 Sep;16(3):297-334. [doi: 10.1007/bf02310555]
33. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med 2016 Jun;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
34. Fayers P, Machin D. Quality of Life: The Assessment, Analysis and Reporting of Patient-reported Outcomes 3rd Edition. United States: Wiley-Blackwell; 2016.
35. Eremenco S, Pease S, Mann S, Berry P, PRO Consortium's Process Subcommittee. Patient-Reported Outcome (PRO) Consortium translation process: consensus development of updated best practices. J Patient Rep Outcomes 2017 Feb 27;2(1):12 [FREE Full text] [doi: 10.1186/s41687-018-0037-6] [Medline: 29757299]
36. Takashina H, Kengo Y. Translation: Review of an app for psychological interventions against depressive symptoms in Japan. PsyArXiv. 2020 July 2020 Jul 09:9. [doi: 10.31234/osf.io/p9m3q]
37. Terwee CB, Bot SDM, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007 Jan;60(1):34-42. [doi: 10.1016/j.jclinepi.2006.03.012] [Medline: 17161752]
38. Kuehnhausen M, Frost V. Trusting smartphone apps? To install or not to install, that is the question. 2013 Presented at: IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA); Feb 25-28, 2013; San Diego, CA. [doi: 10.1109/cogsima.2013.6523820]

Abbreviations

ICC: intraclass correlation coefficient
MARS: Mobile App Rating Scale
mHealth: mobile health

Edited by L Buis, A Mavragani; submitted 22.09.21; peer-reviewed by S Oishi, F Kates, R Romero, RM Payo; comments to author 06.11.21; revised version received 19.12.21; accepted 17.02.22; published 14.04.22.

Please cite as:
Yamamoto K, Ito M, Sakata M, Koizumi S, Hashisako M, Sato M, Stoyanov SR, Furukawa TA
Japanese Version of the Mobile App Rating Scale (MARS): Development and Validation
JMIR Mhealth Uhealth 2022;10(4):e33725
URL: https://mhealth.jmir.org/2022/4/e33725
doi:10.2196/33725
PMID:35197241

Original Paper

Mobile Health Apps Providing Information on Drugs for Adult Emergency Care: Systematic Search on App Stores and Content Analysis

Sebastián García-Sánchez[1*], PharmD, PhD; Beatriz Somoza-Fernández[1*], PharmD; Ana de Lorenzo-Pinto[1], PharmD, PhD; Cristina Ortega-Navarro[1], PharmD; Ana Herranz-Alonso[1], PharmD, PhD; María Sanjurjo[1], PharmD

Pharmacy Department, Instituto de Investigación Sanitaria Gregorio Marañón, Hospital General Universitario Gregorio Marañón, Madrid, Spain

[*]these authors contributed equally

Corresponding Author:
Sebastián García-Sánchez, PharmD, PhD
Pharmacy Department, Instituto de Investigación Sanitaria Gregorio Marañón
Hospital General Universitario Gregorio Marañón
Doctor Esquerdo, 46
Madrid, 28007
Spain
Phone: 34 91 586 82 45
Email: sebasmur@hotmail.com

Abstract

Background: Drug-referencing apps are among the most frequently used by emergency health professionals. To date, no study has analyzed the quantity and quality of apps that provide information on emergency drugs.

Objective: This study aimed to identify apps designed to assist emergency professionals in managing drugs and to describe and analyze their characteristics.

Methods: We performed an observational, cross-sectional, descriptive study of apps that provide information on drugs for adult emergency care. The iOS and Android platforms were searched in February 2021. The apps were independently evaluated by 2 hospital clinical pharmacists. We analyzed developer affiliation, cost, updates, user ratings, and number of downloads. We also evaluated the main topic (emergency drugs or emergency medicine), the number of drugs described, the inclusion of bibliographic references, and the presence of the following drug information: commercial presentations, usual dosage, dose adjustment for renal failure, mechanism of action, therapeutic indications, contraindications, interactions with other medicinal products, use in pregnancy and breastfeeding, adverse reactions, method of preparation and administration, stability data, incompatibilities, identification of high-alert medications, positioning in treatment algorithms, information about medication reconciliation, and cost.

Results: Overall, 49 apps were identified. Of these 49 apps, 32 (65%) were found on both digital platforms; 11 (22%) were available only for Android, and 6 (12%) were available only for iOS. In total, 41% (20/49) of the apps required payment (ranging from €0.59 [US $0.64] to €179.99 [US $196.10]) and 22% (11/49) of the apps were developed by non–health care professionals. The mean weighted user rating was 4.023 of 5 (SD 0.71). Overall, 45% (22/49) of the apps focused on emergency drugs, and 55% (27/49) focused on emergency medicine. More than half (29/47, 62%) did not include bibliographic references or had not been updated for more than a year (29/49, 59%). The median number of drugs was 66 (range 4 to >5000). Contraindications (26/47, 55%) and adverse reactions (24/47, 51%) were found in only half of the apps. Less than half of the apps addressed dose adjustment for renal failure (15/47, 32%), interactions (10/47, 21%), and use during pregnancy and breastfeeding (15/47, 32%). Only 6% (3/47) identified high-alert medications, and 2% (1/47) included information about medication reconciliation. Health-related developer, main topic, and greater amount of drug information were not statistically associated with higher user ratings (P=.99, P=.09, and P=.31, respectively).

Conclusions: We provide a comprehensive review of apps with information on emergency drugs for adults. Information on authorship, drug characteristics, and bibliographic references is frequently scarce; therefore, we propose recommendations to consider when developing an app of these characteristics. Future efforts should be made to increase the regulation of drug-referencing apps and to conduct a more frequent and documented review of their clinical content.

(JMIR Mhealth Uhealth 2022;10(4):e29985) doi:10.2196/29985

KEYWORDS

emergency drugs; emergency medicine; emergency departments; emergency professionals; medication errors; drug characteristics; drug management; apps; mHealth; mobile health; digital health; smartphone; mobile phone

Introduction

Background

Digital technologies are an increasingly relevant resource for health services because they can improve the quality, efficiency, and safety of health care, a particularly relevant issue in the event of emergencies, disasters, and other unplanned care situations [1]. In recent years, there has been a significant increase in the quantity and quality of mobile health apps owing to the efforts made by health professionals and app developers. At the beginning of 2021, almost 50,000 medical apps were available on the main download platforms (Apple App Store and Google Play Store) [2]. Mobile apps are changing the health care landscape because they facilitate the exchange of information among professionals, researchers, and patients and enable easy access to quality services during clinical practice [3,4].

The need for a quick response is one of the most prominent characteristics of emergency medicine. Examples of the high care burden experienced in emergency departments can be seen in the nearly 130 million visits in 2018 in the United States or the 30 million visits registered each year in Spain [5,6]. A variety of apps have been developed in recent years to improve patient care in these departments [7,8]. Medical emergency apps are now a key element of clinical practice as they can be used as clinical decision tools, case management tools, and sources of clinical information. A desirable feature of these apps is that they can be used quickly because of the need to provide a rapid response to the broad spectrum of clinical scenarios occurring in emergency departments. Recent studies on mobile devices and medical apps in emergency rooms [9,10] have shown that the apps most frequently used by emergency health professionals are medical formulary and drug-referencing apps (84.4%), followed by disease diagnosis and management apps (69.5%) [10].

Health care pressure, stressful situations, and the need for multiple high-alert medications make emergency departments the perfect setting for drug-related problems [7]. Insufficient information on drugs is the most common cause of medication errors, which can lead to adverse drug events involving temporary or permanent harm to patients and higher health care costs [11,12]. The information needed in an emergency department includes multiple drug characteristics such as indications, dosing, administration, pharmaceutical compatibilities, adverse reactions, interactions, and contraindications [11,13]. The usefulness of medical apps as a source of information on drug-related characteristics should be highlighted, although the literature still contains relevant gaps concerning these tools. To date, no study has addressed the quantity and quality of smartphone apps that provide information on emergency drugs.

Objective

Therefore, the main objective of this study was to identify apps designed to assist health care professionals in managing drugs for adult emergency care and describe their main characteristics and functionalities. As secondary objectives, we designed a score to estimate the amount of drug information contained in each app and analyzed the relationship between this score and the relevant app characteristics. We also analyzed whether some of the variables selected could affect user satisfaction (app user ratings).

Methods

Search Strategy and App Selection

We performed an observational, cross-sectional, descriptive study of smartphone apps available on the iOS and Android platforms that provide information on drugs used for adult emergency care.

The methodology used for app selection was based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) system [14]. To identify emergency drugs–related apps, a search was conducted between February 15 and February 19, 2021, on the digital distribution platforms Google Play Store (Android) and Apple App Store (iOS), which are the app stores with the most apps available at present [15]. The search terms were “emergency drugs” OR “fármacos de urgencias” and “emergency medicine” OR “medicina de urgencias y emergencias.” We extracted text from app store descriptions and selected apps available in English or Spanish whose content was fully dedicated to drugs commonly used in the emergency room (hereafter referred to as *emergency drugs apps*) and apps related to the field of emergency medicine that contained a section on medications (*emergency medicine apps*). Apps aimed at pediatric emergencies were excluded because of relevant differences in the use of drugs in children (eg, dosage, treatment algorithms, and selection). Both free and paid apps were included. Apps from the Google Play Store were downloaded onto a Xiaomi Mi 9 SE (version 9 PKQ1.181121.001; Android), and apps from the Apple App Store were downloaded onto an iPhone 11 (version 14.4; iOS).

Ethical Considerations

No patients were involved in the study and therefore ethical board approval was not sought, as it is considered unnecessary under RD 1090/2015 regulating clinical trials with medicinal products and the Ethics Committees for Research with medicinal products, and Law 14/2007 on Biomedical Research.

Data Extraction

We collected the following information from the download platforms: app name, operating system (Android, iOS, or both), developer affiliation, country of origin, language, category, cost,

publication date, date of last update, size, version, number of downloads, and user ratings. These indicators are commonly used in studies on health-related apps [16-19]. The overall mean weighted user rating was calculated by considering the number of ratings from both app stores. For the rest of the analysis, when the same app was available on both platforms, we only considered the version available on the Google Play Store as Android is the leading operating system worldwide and the Apple App Store provides less information (no data on the number of downloads). Subsequently, all apps were downloaded and their contents were evaluated. We counted the number of drugs included in each app and determined whether they belonged to ≥1 drug classes. We then evaluated whether the apps contained information on the following fifteen drug-related characteristics: (1) commercial presentations, (2) usual dosage, (3) dose adjustment for renal failure, (4) mechanism of action, (5) therapeutic indications, (6) contraindications, (7) interaction with other medicinal products, (8) use in pregnancy and breastfeeding, (9) adverse reactions, (10) method of preparation and administration, (11) stability data and incompatibilities, (12) identification of high-alert medications, (13) positioning in treatment algorithms, (14) information about medication reconciliation, and (15) cost. The selection of these indicators was discussed by the research team based on the most frequent requests received from emergency medicine pharmacy services and drug information centers [12,13]. High-alert medications are defined as drugs that bear a heightened risk of causing significant patient harm when used erroneously [20]. Medication reconciliation is defined as the formal process in which health care professionals partner with patients to ensure accurate and complete medication information transfer at transitions of care [21].

We assigned a score of 0 to 15 according to the amount of drug information provided in the app. A score of 0 indicates that the app did not include any information about the 15 drug characteristics analyzed, and a score of 15 indicates that all characteristics were shown in the app. Finally, we also evaluated whether the apps included bibliographic references on drug-related concerns.

Data Analysis

All apps were independently evaluated by 2 hospital clinical pharmacists (SGS and BSF). The variables were coded and entered in a Microsoft Excel spreadsheet. The Cohen κ coefficient was calculated using Reliability Calculator for 2 coders [22] to analyze the level of agreement between the data collected by each investigator. Following this analysis, disagreements on the reported results were resolved through iterative discussion and consensus.

A statistical analysis was performed using Stata (version IC-16; StataCorp). On the basis of previously published studies on mobile health apps, we measured the association between a series of app characteristics (developer, main topic, cost, and number of downloads) and user ratings (which indicate user satisfaction) or the score assigned to the app (which indicates the variety of content on drug information). We also analyzed whether the inclusion of bibliographic references could be influenced by the app developer (health-related or non–health-related). The Shapiro-Wilk test was used to evaluate whether continuous variables were normally distributed. For normally distributed data, differences were assessed using the 2-tailed Student *t* test for 2 categories and ANOVA for ≥2 categories; for nonnormally distributed data, the Mann-Whitney *U* test was used. The correlation between quantitative variables was evaluated using the Spearman correlation test. Categorical variables were compared using an uncorrected chi-square test or Fisher exact test, as appropriate. Statistical significance was set at $P<.05$.

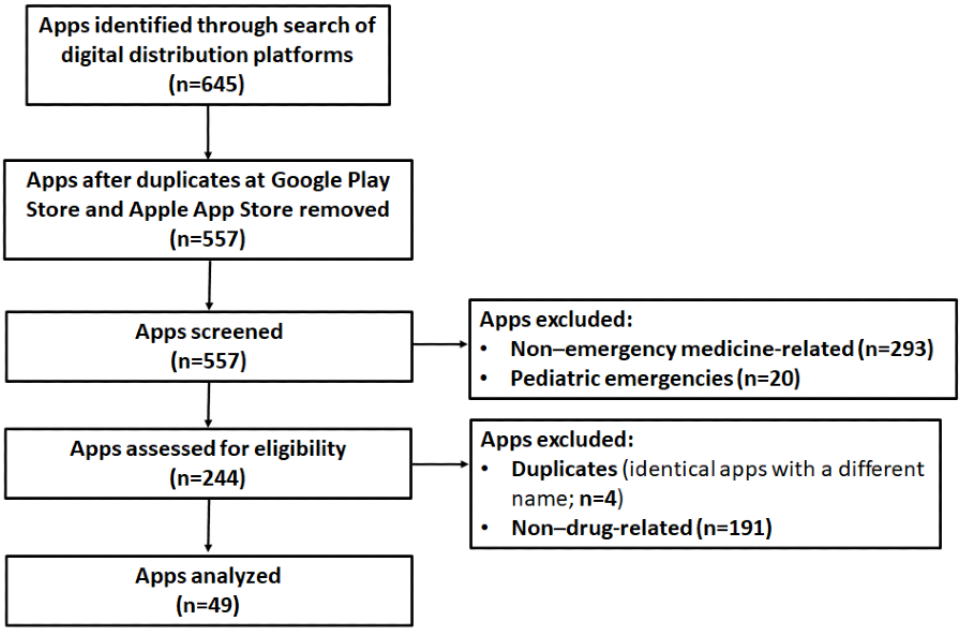
Results

Mobile App Search

Combined keyword searches of the Google Play Store and Apple App Store yielded 645 apps potentially related to emergency drugs. A flow diagram illustrating the selection and exclusion of apps at various stages of the study is shown in Figure 1. We removed 88 apps as duplicates, with the same app name and developer appearing on both download platforms. The remaining 557 apps were further screened. We extracted information from the store app description and removed 293 apps that were not related to emergency medicine and 20 apps aimed at pediatric emergencies. We then exhaustively analyzed the descriptions of the remaining apps and downloaded them to determine whether the information was inaccurate. From the resulting apps, we removed 4 duplicates with different app names within the same store. We eventually excluded 191 apps that did not contain a specific section on drugs. Following this systematic search, we identified 49 apps that met the inclusion criteria. In total, 65% (32/49) of the apps were found on both digital distribution platforms, whereas 22% (11/49) were obtained only from the Google Play Store, and 12% (6/49) were only available from the Apple App Store.

Two independent researchers (SGS and BSF) further analyzed the characteristics, functionalities, and contents of the 49 apps selected. The mean Cohen κ coefficient for interrater reliability was 0.94 (SD 0.05).

Figure 1. A flow diagram illustrating the search process for the apps analyzed in the study.



Analysis of General Characteristics of Apps

Textbox 1 shows the names of the 49 apps classified by their main topic (emergency drugs or emergency medicine).

By origin, 41% (20/49) of the apps were developed in North America, 35% (17/49) in Europe, 12% (6/49) in South America, 4% (2/49) in Asia, and 2% (1/49) in Africa. The origin of 6% (3/49) of the apps could not be determined. Of the 49 apps analyzed, 27 (55%) were published only in English, 21 (43%) were published only in Spanish, and 1 (2%) was available in both languages. Most apps (44/49, 90%) were classified in the category of medicine. The other categories were health and well-being (3/49, 6%) and education (2/49, 4%).

Slightly more than half of the apps were free to download (29/49, 59%), whereas the other 41% (20/49) required payment, with a cost ranging from €0.59 (US $0.64) to €179.99 (US $196.10) (median €8.99 [US $9.79]) and a mean cost of €20.82 (US $22.68) (SD €40.81 [US $44.46]). Two apps were for the exclusive use of workers at the center where they were developed, and 1 app could only be used with a code acquired after purchasing a book; therefore, they could not be fully analyzed. In addition, the content of 1 app was unavailable because of a download error that affected the latest versions of Android. In these cases, we collected as much information as possible from the description of the app and images available on the digital distribution platforms.

The average size of the apps was 23.89 (SD 23.28) MB. The content of 27% (13/49) of the apps was updated 6 months before the search. A further 14% (7/49) of the apps were updated in the previous year. A total of 59% (29/49) of the apps had not been updated for more than a year; of these, 12 (24% of the overall apps) had not been updated for more than 3 years. A total of 16% (8/49) apps had not been updated since the date of the first publication. The average time between the date of analysis and the date of the most recent update was 23.3 (SD 23.6) months. iOS apps were excluded from this last analysis because of the lack of information on the day of the most recent update.

About half of the apps were developed by private and for-profit organizations (22/49, 45%) as follows: health-related technology companies (n=12, 24%); non–health-related technology companies (n=9, 18%); and medical publishers (n=1, 2%). A total of 22% (11/49) apps were developed by non–health-related professionals. Among the 78% (38/49) apps developed by health care professionals, 29% (14/49) were developed by individual professionals, whereas the rest were developed by technology companies or medical publishers (13/49, 27%), or with the involvement of a health care organization (eg, hospital, public health agency, or professional society; 11/49, 22%). A complete list of developers is provided in Table 1.

The number of downloads can only be determined in the apps found in the Google Play Store, as this information is not available in the Apple App Store. The median number of downloads was >5000 (range >1 to >100,000). Detailed information regarding the number of downloads is presented in Table 2.

We evaluated the association between the cost of apps and the number of downloads. Owing to the small sample sizes, the number of downloads was broken down for this analysis into 3 categories: 1 to 1000, 1001 to 10,000, and >10,000 downloads. No statistically significant differences were found between the groups (F_{42}=0.24; P=.70).

The analyses of user ratings included 40 apps, as no data were available for 9 apps. The mean overall weighted user rating of apps according to the number of valuations was 4.023 out of 5 (SD 0.71). The average user ratings were almost identical (t_{38}=−0.01; P=.99) for apps developed by health professionals (n=30, mean 4.240, SD 0.707) and non–health professionals (n=10, mean 4.243, SD 0.470). Free apps were rated higher (n=27, mean 4.277, SD 0.680) than paid apps (n=13, mean 4.197, SD 0.621; t_{38}=−2.27; P=.03).

Textbox 1. List of emergency drugs apps and emergency medicine apps.

Emergency drugs apps

- 50 Drugs in emergency
- Antídotos
- Common 50 drugs for emergency
- Drogas en emergencia y UCI
- ED drugs
- Emergency drugs
- Emergency drugs (Antonio Frontera)
- Emergency drugs (Ferrazza)
- Emergency medication reference
- EMS calculator or EMS drugs fast
- EMS drug cards
- Farmacos de urgencias SES or urgencias SES
- FarmaPoniente
- Goteo para vasoactivos
- Guía farmacológica
- Guía URG
- Infusiones
- Medicina de urgencias
- Paramedic drug list
- Perfusiones urgencias
- Pocket drug guide EMS or EMS pocket guide
- UrgRedFasterFH

Emergency medicine apps

- AHS EMS MedicalProtocols
- Arritmias urgencias
- Basic emergency care
- Chuletario urgencias extrahospitalarias
- EMAT app
- Emergency central
- Emergency medicine on call
- EMR guide
- EMRA antibiotic guide
- EMRA PressorDex
- EMS ACLS guide
- EMS notes: EMT and paramedic
- EMS pro
- Erres
- ICU
- ICU ER facts made Incred quick
- iTox Urgencias intoxicación
- Manual de procedimientos SAMUR

- Médico de urgencias
- My Emergency Department
- Odonto emergencias
- QuickEM
- URG
- Urgencia HBLT or Guia urgencia HBLT
- Urgencias Extrahospitalarias
- WikEM—Medicina de emergencia
- Zubirán. Manual Terapéutica 7e

Table 1. Developers of the apps (N=49).

Developer	Value, n (%)
Individual health professional	14 (29)
Health-related technology company	12 (24)
Non–health-related technology company	9 (18)
Hospital	3 (6)
Public health agency	3 (6)
Individual non–health professional	2 (4)
Medical or pharmaceutical society	2 (4)
Other health professional organization	2 (4)
University	1 (2)
Medical publisher	1 (2)

Table 2. Apps classified by the number of downloads (N=43).

Number of downloads	Value, n (%)
1-100	2 (5)
101-1000	5 (12)
1001-5000	12 (28)
5001-10,000	6 (14)
10,001-100,000	12 (28)
>100,000	6 (14)

Analysis of Contents of Apps

Approximately half of the apps focused on emergency drugs (22/49, 45%), whereas the rest (27/49, 55%) focused on emergency medicine in a broader sense. We did not find statistically significant differences ($z=-1.7$; $P=.09$) between the average user rating of *emergency drugs apps* (4.163/5; 16 apps) and *emergency medicine apps* (4.296/5; 24 apps).

The median number of drugs included in the apps was 66 (range 4 to >5000). The apps classified according to the number of drugs analyzed are shown in Table 3.

Of 49 apps, 6 (12%) analyzed only a specific class of drugs: antidotes (n=2, 33%), vasopressors (n=2, 33%), antibiotics (n=1, 17%), and antiarrhythmics (n=1, 17%).

Table 3. Apps classified by number of drugs analyzed (N=47).

Number of drugs	Value, n (%)
1-25	9 (19)
26-50	12 (26)
51-100	10 (21)
101-200	11 (23)
>200	3 (6)
>1000	2 (4)

Table 4 shows the 15 drug characteristics of the apps analyzed. Most apps included information about therapeutic indications (38/48, 79%) and the most common doses (43/49, 88%). Other drug-related concerns found in more than half of the apps were commercial presentations (27/47, 57%), mechanism of action (26/47, 55%), contraindications (26/47, 55%), method of preparation and administration (25/48, 52%), and adverse reactions (24/47, 51%). Only 17% (8/47) of apps provided data on stability and incompatibilities. Identification of high-alert medications was found in 6% (3/47) of the apps. Information on drug costs was present in only 2% (1/47) of the apps. Similarly, information about medication reconciliation in the emergency room was found in only 2% (1/47) of the apps.

Most apps (29/47, 62%) did not include bibliographic references regarding drug-related concerns. The percentage of apps that included this kind of information was 44% (16/36) in the group of apps developed by health professionals and 18% (2/11) in the group of apps developed by non–health professionals (χ^2_1=2.5; *P*=.12).

Table 4. Drug characteristics described in the apps.

Drug characteristic	Value, N	Value, n (%)
Commercial presentations	47	27 (57)
Usual dosage	49	43 (88)
Dose adjustment for renal failure	47	15 (32)
Mechanism of action	47	26 (55)
Therapeutic indications	48	38 (79)
Contraindications	47	26 (55)
Interaction with other medicinal products	47	10 (21)
Use in pregnancy and breastfeeding	47	15 (32)
Adverse reactions	47	24 (51)
Method of preparation and administration	48	25 (52)
Stability data and incompatibilities	47	8 (17)
Identification of high-alert medications	47	3 (6)
Positioning in treatment algorithms	47	19 (40)
Information about reconciliation	47	1 (2)
Cost	47	1 (2)

Analysis of Drug Information Score

We assigned a score of 0 to 15 according to the number of drug characteristics provided in the app. The mean score was 5.89 (SD 2.91). Of the 47 apps, 22 (47%) apps received a score ranging from 0 to 5, a total of 21 (45%) apps received a score from 6 to 10, and 4 (8%) apps received a score from 11 to 13. There was no correlation between this score and the app user ratings (ρ=−0.17; *P*=.31).

The average score for apps developed by health professionals (n=36, mean 6.00, SD 3.04) was slightly higher than that for apps developed by non–health professionals (n=11, mean 5.55, SD 2.54), although the difference was not significant (t_{45}=−0.45; *P*=.66). Similarly, no statistically significant differences (t_{45}=−0.78; *P*=.44) were found between the average score of emergency drugs apps (n=21, mean 5.52, SD 2.75) and emergency medicine apps (n=26, mean 6.19, SD 3.06) or between the average score of free (n=27, mean 5.90, SD 2.91) and paid (n=20, 5.47, SD 3.07) apps (t_{45}=−0.83; *P*=.40).

Finally, we compared the difference between the number of downloads and drug information score. The average score was 4.57 (SD 1.81) for apps with 1 to 1000 downloads (7/41, 17%), 6.69 (SD 3.28) for apps with 1001 to 10,000 downloads (16/41, 39%), and 6.61 (SD 2.55) for apps with >10,000 downloads

(18/41, 44%). No statistically significant differences were found between the groups (F_{40}=1.63; P=.21).

Discussion

Overview

Studies on the content of mobile health apps are increasingly frequent, and apps related to relevant diseases such as cancer or COVID-19 infection have recently been analyzed [16,17,23,24]. Nevertheless, research on apps designed for use in emergency rooms remains insufficient. In this study, we provide a comprehensive and unique review of smartphone apps that provide information on drugs for adult emergency care.

The use of mobile devices by emergency health professionals is common, and apps related to this field of medicine are proliferating [10,25]. Emergency rooms are areas where a high volume of patients must be seen within a short period, and work interruptions are very frequent [26]. In this complex environment, incorrect use of mobile devices can increase the risk of distraction and may affect patient safety [9]. Nevertheless, when these devices are used properly, they have enormous potential to improve medical practice, for instance, by allowing quick access to relevant and evidence-based information, which facilitates decision-making and can help reduce error rates. In a recent survey of professionals in an emergency department, most respondents found mobile devices useful for better coordinating care among providers and beneficial for patient care [10].

Principal Findings on General Characteristics and Comparison With Prior Studies

Our study provides a general perspective on apps designed to help health care professionals with drug management for adult emergency care. Given that medication errors are commonly caused by insufficient information on drugs [12], we analyzed these apps in detail. This is one of the most comprehensive studies of apps aimed at providing information about drugs for health care professionals. Recently, a study identified more than 600 drug-related apps, and approximately two-third of them were categorized within the medication information class [27]. The authors distinguished among apps for patients, apps for health professionals, and apps that can be used by both groups. Recent studies on patient-focused drug apps have analyzed those that help patients understand and take their medications or those with a medication list function [28,29]. In addition, apps for treatment adherence have been the subject of intensive research [30-35]. Some papers have also been published on apps about drug-drug interactions [36,37]. This is an issue traditionally addressed by health care professionals, although nowadays many apps for checking interactions are intended to be used by patients rather than health care professionals.

Knowledge of the characteristics of drug apps designed to be used exclusively by health care professionals is still limited. Few studies have aimed to analyze the functionalities and content of these apps. A study conducted in 2013 identified 306 apps providing drug reference information and prescribing material, and analyzed cost, updates, user ratings, intended area of use, and medical involvement in app development [38]. More recently, a study published in 2017 compared 8 apps for dosage recommendations, adverse reactions, and drug interactions [39]. The quality of the apps targeting medication-related problems has been assessed. Of the 59 apps analyzed, 23 (39%) contained medication information features [40]. Very recently, a study identified 23 drug reference apps with local drug information in Taiwan (including those aimed at both patients and professionals) and analyzed their quality and factors influencing user perceptions [41]. In the field of emergency care, a recent study analyzed apps for the management of drug poisoning [42]. Of the 17 apps identified, 14 (82%) presented diagnosis and treatment guides, and 3 (18%) were specifically on antidotes and their dosage.

In our study, we first collected the information available in the app marketplace descriptions (eg, number of downloads and user ratings) before downloading the apps and analyzing their content in detail. This strategy differs from those of other recent studies, in which a greater number of apps were identified but where the analysis was limited to the marketplace description [8,18,19,38]. Among our main findings, we can highlight that 22% (11/49) of the apps were not developed by health care professionals. This is a lower percentage than that reported in other studies on mobile health apps [18,23,43]. In 2013, there was no evidence of involvement of health care professionals in the development of 32.7% (100/306) of the apps available to support prescribing practice [38]. It should also be noted that apps for patient medication management are developed mainly by the software industry, without the involvement of health care professionals [28,31]. Nevertheless, our findings should be considered relevant, given that the apps we analyzed are intended to be used in complex and emergency situations. In addition, information on authorship is scarce in many of the apps evaluated.

More importantly, we found that more than half of the apps (29/47, 62%) did not include bibliographic references or had not been updated for more than a year (29/49, 59%). Our results are in accordance with a previous study analyzing 23 apps with medication information, most of which did not provide supporting references [40]. Of particular concern is the lack of updates in the apps analyzed in our study, as this indicator has worsened compared with the study conducted by Haffey et al [38], in which 44.4% (136/306) of the apps had either been released or updated within the last 6 months, and a further 24.2% (74/306) within 1 year [38]. These concerns raise doubts about the quality and reliability of the information provided by these apps aimed at emergency health care professionals, as incorrect drug information may remain for long periods.

Doubts arise when a health app is developed by non–health professionals [44,45]. We found that bibliographic references were included in 44% (16/36) of the apps developed by health professionals and in only 18% (2/11) of the apps developed by non–health professionals. This result was not statistically significant (P=.12), probably because of the small sample size, although it highlights the uncertainty surrounding the sources of information provided in apps developed by non–health professionals. The reliability and authority of information should be analyzed by health care professionals who are more capable of evaluating, reviewing, and verifying the content of

health-related apps. In the field of medication, pharmacists should play a vital role in reviewing apps.

About half of the apps (20/49, 41%) required payment to access all the content, with a cost ranging from €0.59 (US $0.64) to €179.99 (US $196.10). This is a similar percentage than that observed in a recent study on drug poisoning management apps [42]. Nevertheless, it is considerably higher than that observed in other recent reviews of apps for medical emergencies [8], medication management and adherence for patients [28,32], or checking for drug-drug interactions [36]. We hypothesize that these differences could arise because the apps analyzed in our study are aimed exclusively at health care professionals and are designed for use in health care facilities. A study conducted in 2013 on apps to support drug prescribing or provide pharmacology education showed that 68% (208/306) of the apps required payment, with a mean price of £14.25 (US $18.57) per app and a range of £0.62 (US $0.81) to £101.90 (US $132.76) [38]. The cost of apps also seems to be influenced by the origin of the developer [41]. In any case, cost is an important determinant in the decision to adopt a mobile health app, regardless of age group and socioeconomic status [46]. In addition, payment for the apps analyzed in our study could be a relevant limitation for health care professionals who only occasionally work in emergency rooms, as is common in many hospitals.

To date, few studies have analyzed the factors that influence user satisfaction with apps [18,47,48]. The number of downloads and user ratings are usually correlated and have been proposed as indicators of acceptability and satisfaction with mobile health apps [49,50]. A secondary objective of our study was to learn more about user behavior with emergency medicine apps, for which we analyzed whether factors such as cost, the main theme of the app (emergency medicine or emergency drugs), or the app developer (health-related or non–health related) could influence user ratings. The free apps analyzed in our study had higher user ratings than paid apps, although no association was found between the cost and number of downloads. We found no further statistically significant differences, probably because of the small sample size. The number of downloads and user ratings probably depend on multiple factors. Navigation, performance, visual appeal, credibility, and quantity of information have recently been identified as the most influential factors on higher user ratings in a study analyzing 23 drug reference apps [41]. Previous studies have reported highly variable results for the influence of expert involvement in app development on user ratings and the number of downloads [18,41,49]. In any case, user ratings and downloads should not be considered good predictors of the quality and reliability of medical apps because they could be influenced by other factors, such as low price, in-app purchase options, in-app advertisements, and recent updates [18,51,52]. In our study, the number of downloads, cost, and user ratings were not associated with a score created to quantify the variety of relevant information on drug characteristics in the apps (P=.21, P=.40, and P=.31, respectively). Further research should analyze the reliability of the clinical content of drug information apps and corroborate its association with a greater intention to use or better user satisfaction.

Drug Information Gaps

At present, there are no standardized guidelines for assessing the clinical content and quality of mobile health apps [18]. A highly specific quality assessment tool was developed to assess the quality of apps targeting medication-related problems, including those with medication information features [40]. Nevertheless, the most commonly used methodology to assess the quality of medical apps is the Mobile Application Rating Scale [53-55], as well as in studies on drug apps [30,36,37,41]. The total number of features has been associated with the total Mobile Application Rating Scale score in a study on apps for potential drug-drug interaction decision support [36].

In our study, we paid special attention to the information provided by the apps regarding relevant drug characteristics. We found that most apps included information about the usual dosage (43/49, 88%) and therapeutic indications (38/48, 79%). Nevertheless, other relevant characteristics were found in less than half of the apps, such as dose adjustment for renal failure (15/47, 32%) and use in pregnancy and breastfeeding (15/47, 32%). Interaction with other medicinal products was found in only 21% (10/47) of the apps, despite being a major problem in patient safety. Drug-drug interaction checks are one of the most frequent functional categories within the current medication-related app landscape [27,56], but relevant quality and accuracy problems have been detected in apps, including this feature [36,37]. In addition, other relevant information on drug safety, such as contraindications (26/47, 55%) and adverse reactions (24/47, 51%), was found in approximately half of the apps analyzed in our study. Furthermore, it is worrying that only 6% (3/47) of the apps clearly identified high-alert medications, despite efforts made to avoid errors with these drugs [12].

In clinical practice, many medication-related inquiries are about the method of administration; however, our study showed that this information is included in slightly more than half of the apps (25/48, 52%). In addition, stability data and incompatibilities were present in only 17% (8/47) of the apps. Nurses have also been reported to be frequent app users in daily practice, albeit at a slightly lower percentage than that observed by physicians [10]. Therefore, apps for the use of drugs in the emergency department should be designed to provide more information on drug administration characteristics.

Finally, incorrect medication reconciliation in the emergency department can lead to relevant medication errors [57]. We found only 1 app that appropriately addressed this issue, including information on the maximum time to carry out reconciliation or the possible presence of withdrawal syndrome. Given that medication reconciliation has been considered the most relevant activity carried out by pharmacists in emergency departments [58], it would be desirable for apps related to emergency drugs to provide more information on this matter.

Recommendations for Development of an Emergency Drugs App

There are a growing number of health apps on the market with highly variable designs and content, and it is difficult to determine which are the most useful for health care

professionals. Given the relative absence of legislation on medical apps [59] and the risks associated with drugs used in the emergency room, it would be interesting to propose a series of improvements in the content of apps for emergency drugs. The results of our study and clinical experience enable us to make several recommendations.

Design, ease of use, and the ability to quickly respond to questions that arise during daily clinical practice are especially relevant characteristics, considering that these apps are to be used in a stressful environment. The success of an app for emergency professionals depends on quickly obtaining a reliable response.

Our findings could help developers design apps that provide drug-related information most frequently demanded by health care professionals. Drug information centers have historically received the most inquiries regarding therapeutic indications, adverse reactions, and identification of medical products [13]. In addition, information on contraindications, appropriate dosage, and major drug-drug interactions should be included to prevent major adverse events [11]. We provided a score to measure the amount of drug information included in each app, and our results showed that a greater amount of information is not necessarily associated with better user ratings. Therefore, it could be beneficial to design apps with content aimed exclusively at doctors and apps for nurses, although with maximum information of interest for each of these professionals. For example, apps with information on drug administration and incompatibilities would have the potential to help nursing staff by reducing their workload and, ultimately, the risk of drug-related errors. Strategies to identify high-alert medications should be included in all emergency drug apps, regardless of the group of health care professionals they focus on [12].

We recommend caution with respect to the sources of information used to elaborate the content of the app to ensure that it is reliable. Apps should only be considered reliable based on an extensive literature review, expert panel review, or peer review. The author's affiliation and bibliographic references to scientific and clinical evidence should always be clearly shown [60], and health professionals participating in reviewing and app updates should be clearly identified.

As previously suggested [16], we believe that future legislation should require a more comprehensive description of the mobile app marketplace, with detailed information on authorship and the process used to review app functionalities and the clinical information provided. All information must be supported by appropriate bibliographic references, and developers should preferably be clinicians with experience writing or synthesizing medical evidence. Thus, the information provided, which should be checked by independent reviewers or endorsed by health organizations of recognized prestige, will be more reliable. In addition, we suggest that app developers clearly identify the target user group and provide the maximum amount of drug information relevant to each professional category. It may also be relevant for a partner with a technology company to make apps more attractive and user-friendly.

Limitations and Future Research Directions

First, our study was limited by the inclusion criteria. There are hundreds (perhaps thousands) of apps providing drug-related information, and some may be useful for emergency room professionals; however, they were not analyzed in this study because our aim was to review apps specifically related to emergency drugs or medicine in adults. Drug information indicators were selected and analyzed by the authors and were therefore not validated. A more comprehensive analysis of drug information apps may be the subject of future research, for which our methodology could prove useful. Our approach could be adapted to analyze apps related to child health care or to include indicators not described in our study, such as information on pharmacokinetic properties, therapeutic drug monitoring, and pharmacogenomics. Other limitations are associated with the study design. We only analyzed the Android version when the same app was available on Android and iOS platforms. It should be noted that some characteristics, such as the date of the last update, may vary among platforms. In addition, we analyzed apps in English and Spanish. Although Spanish is the language with the second highest number of native speakers, many health professionals are not sufficiently competent in the language to use these apps comfortably. Our study was also limited by the fact that it only analyzed whether a series of drug characteristics of interest were included in the app. Further research is needed to evaluate the clinical accuracy of the drug information provided by the apps. One possible approach would be a peer-review process to evaluate app contents in terms of the reliability and quality of information according to the best available clinical evidence.

Conclusions

We conducted a comprehensive and unique systematic review of apps that provide information on drugs for adult emergency care. We identified 49 apps according to the PRISMA methodology and conducted a content analysis on most of them. Health-related app developers, the main topic of the app (emergency drugs or emergency medicine), and a greater amount of drug information were not associated with higher app user ratings. Slightly less than half of the apps (20/49, 41%) required payment, with a cost ranging from €0.59 (US $0.64) to €179.99 (US $196.10). We noted that 22% (11/49) of the apps were not developed by health care professionals. Most apps include information about the usual dosage and therapeutic indications, although information on safety and drug administration is much less frequent. Very few apps provide relevant information, such as high-alert medication notices and instructions for drug reconciliation. In addition, more than half of the apps (29/47, 62%) did not include bibliographic references. These findings cast doubts on the quality of many apps. Therefore, we propose a series of issues that should be considered when developing an app of these characteristics and advocate for greater regulation and more frequent and documented review of app content.

Acknowledgments

The authors would like to thank Thomas O'Boyle for editing and proofreading the manuscript.

Authors' Contributions

This study was the result of a collaboration between all the authors. SGS and BSF contributed equally to this work. SGS designed the study, supervised and performed the data collection, and drafted the manuscript. BSF conducted the data analysis and substantially contributed to data collection and drafting of the manuscript. AdLP and CON conceived the original research idea and made a considerable contribution to the design of the study. All authors critically revised the manuscript and approved its final version for publication.

Conflicts of Interest

None declared.

References

1. mHealth: use of appropriate digital technologies for public health: report by the Director-General. World Health Organization. 2017. URL: https://apps.who.int/iris/handle/10665/274134 [accessed 2022-04-01]
2. Number of mHealth apps available at Google Play and App Store from 1st quarter 2015 to 3rd quarter 2020. Statista. URL: https://www.statista.com/statistics/779919/health-apps-available-google-play-worldwide/ [accessed 2022-04-01]
3. Alonso SG, Marques G, Barrachina I, Garcia-Zapirain B, Arambarri J, Salvador JC, et al. Telemedicine and e-Health research solutions in literature for combatting COVID-19: a systematic review. Health Technol (Berl) 2021;11(2):257-266 [FREE Full text] [doi: 10.1007/s12553-021-00529-7] [Medline: 33558838]
4. de la Torre Díez I, Alonso SG, Hamrioui S, López-Coronado M, Cruz EM. Systematic review about QoS and QoE in telemedicine and eHealth services and applications. J Med Syst 2018 Aug 29;42(10):182. [doi: 10.1007/s10916-018-1040-4] [Medline: 30155565]
5. Emergency department visits. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/nchs/fastats/emergency-department.htm [accessed 2022-04-01]
6. Informe Anual del Sistema Nacional de Salud 2018 Resumen ejecutivo. Ministerio de Sanidad. URL: https://www.mscbs.gob.es/estadEstudios/estadisticas/sisInfSanSNS/tablasEstadisticas/InfAnualSNS2018/ResumenEjecutivo2018.pdf [accessed 2022-04-01]
7. Farmer B. Patient safety in the emergency department. Emerg Med 2016 Sep 01;48(9):396-404. [doi: 10.12788/emed.2016.0052]
8. Plaza Roncero A, Marques G, Sainz-De-Abajo B, Martín-Rodríguez F, Del Pozo Vegas C, Garcia-Zapirain B, et al. Mobile health apps for medical emergencies: systematic review. JMIR Mhealth Uhealth 2020 Dec 11;8(12):e18513 [FREE Full text] [doi: 10.2196/18513] [Medline: 33306037]
9. Alameddine M, Soueidan H, Makki M, Tamim H, Hitti E. The use of smart devices by care providers in emergency departments: cross-sectional survey design. JMIR Mhealth Uhealth 2019 May 05;7(6):e13614 [FREE Full text] [doi: 10.2196/13614] [Medline: 31199328]
10. Hitti E, Hadid D, Melki J, Kaddoura R, Alameddine M. Mobile device use among emergency department healthcare professionals: prevalence, utilization and attitudes. Sci Rep 2021 Jan 21;11(1):1917 [FREE Full text] [doi: 10.1038/s41598-021-81278-5] [Medline: 33479264]
11. Hammond DA, Gurnani PK, Flannery AH, Smetana KS, Westrick JC, Lat I, et al. Scoping review of interventions associated with cost avoidance able to be performed in the intensive care unit and emergency department. Pharmacotherapy 2019 Mar;39(3):215-231. [doi: 10.1002/phar.2224] [Medline: 30664269]
12. Ortmann MJ, Johnson EG, Jarrell DH, Bilhimer M, Hayes BD, Mishler A, et al. ASHP guidelines on emergency medicine pharmacist services. Am J Health Syst Pharm 2021 Jan 22;78(3):261-275. [doi: 10.1093/ajhp/zxaa378] [Medline: 33480409]
13. Rosenberg JM, Koumis T, Nathan JP, Cicero LA, McGuire H. Current status of pharmacist-operated drug information centers in the United States. Am J Health Syst Pharm 2004 Oct 01;61(19):2023-2032. [doi: 10.1093/ajhp/61.19.2023] [Medline: 15509125]
14. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64 [FREE Full text] [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
15. Mobile operating system market share worldwide. Statcounter Global Stats. URL: https://gs.statcounter.com/os-market-share/mobile/worldwide [accessed 2022-04-01]
16. Charbonneau DH, Hightower S, Katz A, Zhang K, Abrams J, Senft N, et al. Smartphone apps for cancer: a content analysis of the digital health marketplace. Digit Health 2020;6:2055207620905413 [FREE Full text] [doi: 10.1177/2055207620905413] [Medline: 32110428]

17. Collado-Borrell R, Escudero-Vilaplana V, Villanueva-Bueno C, Herranz-Alonso A, Sanjurjo-Saez M. Features and functionalities of smartphone apps related to COVID-19: systematic search in app stores and content analysis. J Med Internet Res 2020 Aug 25;22(8):e20334 [FREE Full text] [doi: 10.2196/20334] [Medline: 32614777]
18. Biviji R, Vest JR, Dixon BE, Cullen T, Harle CA. Factors related to user ratings and user downloads of mobile apps for maternal and infant health: cross-sectional study. JMIR Mhealth Uhealth 2020 Jan 24;8(1):e15663 [FREE Full text] [doi: 10.2196/15663] [Medline: 32012107]
19. Aydin G, Silahtaroglu G. Insights into mobile health application market via a content analysis of marketplace data with machine learning. PLoS One 2021;16(1):e0244302 [FREE Full text] [doi: 10.1371/journal.pone.0244302] [Medline: 33406100]
20. High-alert medications in acute care settings. Institute for Safe Medication Practices. URL: https://www.ismp.org/recommendations/high-alert-medications-acute-list [accessed 2022-04-01]
21. Assuring medication accuracy at transitions in care: medication reconciliation. The High 5s Project. URL: https://www.who.int/patientsafety/implementation/solutions/high5s/h5s-guide.pdf?ua=1 [accessed 2022-04-01]
22. Freelon D. ReCal: intercoder reliability calculation as a web service. Int J Internet Sci 2010;5(1):20 [FREE Full text]
23. Collado-Borrell R, Escudero-Vilaplana V, Ribed-Sánchez A, Ibáñez-García S, Herranz-Alonso A, Sanjurjo-Sáez M. Smartphone applications for cancer patients; what we know about them? Farm Hosp 2016 Jan 01;40(1):25-35 [FREE Full text] [doi: 10.7399/fh.2016.40.1.8993] [Medline: 26882831]
24. Kondylakis H, Katehakis DG, Kouroubali A, Logothetidis F, Triantafyllidis A, Kalamaras I, et al. COVID-19 mobile apps: a systematic review of the literature. J Med Internet Res 2020 Dec 09;22(12):e23170 [FREE Full text] [doi: 10.2196/23170] [Medline: 33197234]
25. Gaziel-Yablowitz M, Schwartz DG. A review and assessment framework for mobile-based emergency intervention apps. ACM Comput Surv 2019 Jan 31;51(1):1-32. [doi: 10.1145/3145846]
26. Chisholm CD, Dornfeld AM, Nelson DR, Cordell WH. Work interrupted: a comparison of workplace interruptions in emergency departments and primary care offices. Ann Emerg Med 2001 Aug;38(2):146-151. [doi: 10.1067/mem.2001.115440] [Medline: 11468609]
27. Degelsegger-Márquez A, Trunner K, Piso B. Medication apps - a systematic search and classification. Stud Health Technol Inform 2020 Jun 23;271:207-214. [doi: 10.3233/SHTI200098] [Medline: 32578565]
28. Tabi K, Randhawa AS, Choi F, Mithani Z, Albers F, Schnieder M, et al. Mobile apps for medication management: review and analysis. JMIR Mhealth Uhealth 2019 Sep 11;7(9):e13608 [FREE Full text] [doi: 10.2196/13608] [Medline: 31512580]
29. Diaz-Skeete YM, McQuaid D, Akinosun AS, Ekerete I, Carragher N, Carragher L. Analysis of apps with a medication list functionality for older adults with heart failure using the mobile app rating scale and the IMS institute for healthcare informatics functionality score: evaluation study. JMIR Mhealth Uhealth 2021 Nov 02;9(11):e30674 [FREE Full text] [doi: 10.2196/30674] [Medline: 34726613]
30. Santo K, Richtering SS, Chalmers J, Thiagalingam A, Chow CK, Redfern J. Mobile phone apps to improve medication adherence: a systematic stepwise process to identify high-quality apps. JMIR Mhealth Uhealth 2016 Dec 02;4(4):e132 [FREE Full text] [doi: 10.2196/mhealth.6742] [Medline: 27913373]
31. Ahmed I, Ahmad NS, Ali S, Ali S, George A, Saleem Danish H, et al. Medication adherence apps: review and content analysis. JMIR Mhealth Uhealth 2018 Mar 16;6(3):e62 [FREE Full text] [doi: 10.2196/mhealth.6432] [Medline: 29549075]
32. Park JY, Li J, Howren A, Tsao NW, De Vera M. Mobile phone apps targeting medication adherence: quality assessment and content analysis of user reviews. JMIR Mhealth Uhealth 2019 Jan 31;7(1):e11919 [FREE Full text] [doi: 10.2196/11919] [Medline: 30702435]
33. Pérez-Jover V, Sala-González M, Guilabert M, Mira JJ. Mobile apps for increasing treatment adherence: systematic review. J Med Internet Res 2019 Jun 18;21(6):e12505 [FREE Full text] [doi: 10.2196/12505] [Medline: 31215517]
34. Armitage LC, Kassavou A, Sutton S. Do mobile device apps designed to support medication adherence demonstrate efficacy? A systematic review of randomised controlled trials, with meta-analysis. BMJ Open 2020 Jan 30;10(1):e032045 [FREE Full text] [doi: 10.1136/bmjopen-2019-032045] [Medline: 32005778]
35. Al-Arkee S, Mason J, Lane DA, Fabritz L, Chua W, Haque MS, et al. Mobile apps to improve medication adherence in cardiovascular disease: systematic review and meta-analysis. J Med Internet Res 2021 May 25;23(5):e24190 [FREE Full text] [doi: 10.2196/24190] [Medline: 34032583]
36. Kim BY, Sharafoddini A, Tran N, Wen EY, Lee J. Consumer mobile apps for potential drug-drug interaction check: systematic review and content analysis using the mobile app rating scale (MARS). JMIR Mhealth Uhealth 2018 Mar 28;6(3):e74 [FREE Full text] [doi: 10.2196/mhealth.8613] [Medline: 29592848]
37. Shen C, Jiang B, Yang Q, Wang C, Lu KZ, Gu M, et al. Mobile apps for drug-drug interaction checks in chinese app stores: systematic review and content analysis. JMIR Mhealth Uhealth 2021 Jun 15;9(6):e26262 [FREE Full text] [doi: 10.2196/26262] [Medline: 33962910]
38. Haffey F, Brady RR, Maxwell S. Smartphone apps to support hospital prescribing and pharmacology education: a review of current provision. Br J Clin Pharmacol 2014 Jan;77(1):31-38 [FREE Full text] [doi: 10.1111/bcp.12112] [Medline: 23488599]

39. Apidi NA, Murugiah MK, Muthuveloo R, Soh YC, Caruso V, Patel R, et al. Mobile medical applications for dosage recommendation, drug adverse reaction, and drug interaction: review and comparison. Ther Innov Regul Sci 2017 Jul;51(4):480-485. [doi: 10.1177/2168479017696266] [Medline: 30227053]
40. Loy JS, Ali EE, Yap KY. Quality assessment of medical apps that target medication-related problems. J Manag Care Spec Pharm 2016 Oct;22(10):1124-1140. [doi: 10.18553/jmcp.2016.22.10.1124] [Medline: 27668561]
41. Chen Y, Liao W, Su M, Lin Y. Personalized and self-management: systematic search and evaluation quality factors and user preference of drug reference apps in Taiwan. J Pers Med 2021 Aug 12;11(8):790 [FREE Full text] [doi: 10.3390/jpm11080790] [Medline: 34442435]
42. Hajesmaeel Gohari S, Bahaadinbeigy K, Tajoddini S, R Niakan Kalhori S. Drug poisoning management using smartphones: an apps review study based on use case classification model. BMJ Innov 2020 Jul 06;6(4):246-251. [doi: 10.1136/bmjinnov-2020-000447]
43. Subhi Y, Bube SH, Rolskov Bojsen S, Skou Thomsen AS, Konge L. Expert involvement and adherence to medical evidence in medical mobile phone apps: a systematic review. JMIR Mhealth Uhealth 2015 Jul 27;3(3):e79 [FREE Full text] [doi: 10.2196/mhealth.4169] [Medline: 26215371]
44. van Velsen L, Beaujean DJ, van Gemert-Pijnen JE. Why mobile health app overload drives us crazy, and how to restore the sanity. BMC Med Inform Decis Mak 2013 Feb 11;13:23 [FREE Full text] [doi: 10.1186/1472-6947-13-23] [Medline: 23399513]
45. Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mHealth: where is the evidence? PLoS Med 2013;10(2):e1001382 [FREE Full text] [doi: 10.1371/journal.pmed.1001382] [Medline: 23424286]
46. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158 [FREE Full text] [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]
47. Mendiola MF, Kalnicki M, Lindenauer S. Valuable features in mobile health apps for patients and consumers: content analysis of apps and user ratings. JMIR Mhealth Uhealth 2015 May 13;3(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4283] [Medline: 25972309]
48. Carlo AD, Hosseini Ghomi R, Renn BN, Areán PA. By the numbers: ratings and utilization of behavioral health mobile applications. NPJ Digit Med 2019;2:54 [FREE Full text] [doi: 10.1038/s41746-019-0129-6] [Medline: 31304400]
49. Pereira-Azevedo N, Osório L, Cavadas V, Fraga A, Carrasquinho E, Cardoso de Oliveira E, et al. Expert involvement predicts mHealth app downloads: multivariate regression analysis of urology apps. JMIR Mhealth Uhealth 2016 Jul 15;4(3):e86 [FREE Full text] [doi: 10.2196/mhealth.5738] [Medline: 27421338]
50. Huang H, Bashir M. Users' adoption of mental health apps: examining the impact of information cues. JMIR Mhealth Uhealth 2017 Jun 28;5(6):e83 [FREE Full text] [doi: 10.2196/mhealth.6827] [Medline: 28659256]
51. Ghose A, Han SP. Estimating demand for mobile applications in the new economy. Manag Sci 2014 Jun;60(6):1470-1488. [doi: 10.1287/mnsc.2014.1945]
52. Plante TB, O'Kelly AC, Macfarlane ZT, Urrea B, Appel LJ, Miller Iii ER, et al. Trends in user ratings and reviews of a popular yet inaccurate blood pressure-measuring smartphone app. J Am Med Inform Assoc 2018 Aug 01;25(8):1074-1079 [FREE Full text] [doi: 10.1093/jamia/ocy060] [Medline: 29878236]
53. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
54. Amor-García M, Collado-Borrell R, Escudero-Vilaplana V, Melgarejo-Ortuño A, Herranz-Alonso A, Arija A, et al. Assessing apps for patients with genitourinary tumors using the mobile application rating scale (MARS): systematic search in app stores and content analysis. JMIR Mhealth Uhealth 2020 Jul 23;8(7):e17609 [FREE Full text] [doi: 10.2196/17609] [Medline: 32706737]
55. Woulfe F, Fadahunsi KP, Smith S, Chirambo GB, Larsson E, Henn P, et al. Identification and evaluation of methodologies to assess the quality of mobile health apps in high-, low-, and middle-income countries: rapid review. JMIR Mhealth Uhealth 2021 Oct 12;9(10):e28384 [FREE Full text] [doi: 10.2196/28384] [Medline: 34636737]
56. Hammar T, Hamqvist S, Zetterholm M, Jokela P, Ferati M. Current knowledge about providing drug-drug interaction services for patients-a scoping review. Pharmacy (Basel) 2021 Mar 24;9(2):69 [FREE Full text] [doi: 10.3390/pharmacy9020069] [Medline: 33805205]
57. Santell JP. Reconciliation failures lead to medication errors. Jt Comm J Qual Patient Saf 2006 Apr;32(4):225-229. [doi: 10.1016/s1553-7250(06)32029-6] [Medline: 16649654]
58. García-Martín Á, Maroun-Eid C, Campino-Villegas A, Oliva-Manuel B, Herrero-Ambrosio A, Quintana-Díaz M. Perception survey on the value of the hospital pharmacist at the emergency department. Farm Hosp 2017 May 01;41(3):357-370. [doi: 10.7399/fh.2017.41.3.10687] [Medline: 28478750]
59. Gordon WJ, Landman A, Zhang H, Bates DW. Beyond validation: getting health apps into clinical practice. NPJ Digit Med 2020;3:14 [FREE Full text] [doi: 10.1038/s41746-019-0212-z] [Medline: 32047860]
60. Health on the net foundation. HONcode principles. URL: https://www.hon.ch/HONcode/Webmasters/Conduct.html [accessed 2022-04-01]

Abbreviations

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

Edited by L Buis; submitted 27.04.21; peer-reviewed by E Hitti, L Collado-Yurrita, V Cavadas, V Escudero-Vilaplana, H Soueidan; comments to author 22.11.21; revised version received 15.02.22; accepted 18.02.22; published 20.04.22.

Please cite as:
García-Sánchez S, Somoza-Fernández B, de Lorenzo-Pinto A, Ortega-Navarro C, Herranz-Alonso A, Sanjurjo M
Mobile Health Apps Providing Information on Drugs for Adult Emergency Care: Systematic Search on App Stores and Content Analysis
JMIR Mhealth Uhealth 2022;10(4):e29985
URL: https://mhealth.jmir.org/2022/4/e29985
doi:10.2196/29985
PMID:35442212

Original Paper

A Mobile Phone App to Support Adherence to Daily HIV Pre-exposure Prophylaxis Engagement Among Young Men Who Have Sex With Men and Transgender Women Aged 15 to 19 Years in Thailand: Pilot Randomized Controlled Trial

Surinda Kawichai[1], PhD; Wipaporn Natalie Songtaweesin[2], MBBS, DTMH; Prissana Wongharn[1], MSc; Nittaya Phanuphak[3], MD, PhD; Tim R Cressey[4,5], PhD; Juthamanee Moonwong[1], MSc; Anuchit Vasinonta[6], BEng, MEE; Chutima Saisaengjan[1], MA; Tanat Chinbunchorn[3], BSc, MD; Thanyawee Puthanakit[1,7], MD

[1]Center of Excellence for Pediatric Infectious Diseases and Vaccines, Chulalongkorn University, Bangkok, Thailand

[2]Center of Excellence in Transgender Health, Chulalongkorn University, Bangkok, Thailand

[3]Institute of HIV Research and Innovation, Bangkok, Thailand

[4]The Program for HIV Prevention and Treatment/ Unité Mixte de recherches Internationale 174, Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai, Thailand

[5]Department of Molecular and Clinical Pharmacology, University of Liverpool, Liverpool, United Kingdom

[6]Focal Intelligence Co Ltd, Bangkok, Thailand

[7]Department of Pediatrics, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand

Corresponding Author:
Wipaporn Natalie Songtaweesin, MBBS, DTMH
Center of Excellence in Transgender Health
Chulalongkorn University
King Chulalongkorn Memorial Hospital
1873 Rama 4 Rd, Patumwan
Bangkok, 10330
Thailand
Phone: 66 22564930
Email: wipaporn.n@chula.ac.th

Abstract

Background: Widespread smartphone use provides opportunities for mobile health HIV prevention strategies among at-risk populations.

Objective: This study aims to investigate engagement in a theory-based (information–motivation–behavioral skills model) mobile phone app developed to support HIV pre-exposure prophylaxis (PrEP) adherence among Thai young men who have sex with men (YMSM) and young transgender women (YTGW) in Bangkok, Thailand.

Methods: A randomized controlled trial was conducted among HIV-negative YMSM and YTGW aged 15-19 years initiating daily oral PrEP. Participants were randomized to receive either youth-friendly PrEP services (YFS) for 6 months, including monthly contact with site staff (clinic visits or telephone follow-up) and staff consultation access, or YFS plus use of a PrEP adherence support app (YFS+APP). The target population focus group discussion findings and the information–motivation–behavioral skills model informed app development. App features were based on the *3Rs*—risk assessment of self-HIV acquisition risk, reminders to take PrEP, and rewards as redeemable points. Dried blood spots quantifying of tenofovir diphosphate were collected at months 3 and 6 to assess PrEP adherence. Tenofovir diphosphate ≥350-699 fmol/punch was classified as fair adherence and ≥700 fmol/punch as good adherence. Data analysis on app use paradata and exit interviews were conducted on the YFS+APP arm after 6 months of follow-up.

Results: Between March 2018 and June 2019, 200 participants with a median age of 18 (IQR 17-19) years were enrolled. Overall, 74% (148/200) were YMSM; 87% (87/100) of participants who received YFS+APP logged in to the app and performed weekly HIV acquisition risk assessments (log-in and risk assessment [LRA]). The median duration between the first and last log-in was 3.5 (IQR 1.6-5.6) months, with a median frequency of 6 LRAs (IQR 2-10). Moreover, 22% (22/100) of the participants in the YFS+APP arm were frequent users (LRA≥10) during the 6-month follow-up period. YMSM were 9.3 (95% CI 1.2-74.3)

times more likely to be frequent app users than YTGW (*P*=.04). Frequent app users had higher proportions (12%-16%) of PrEP adherence at both months 3 and 6 compared with infrequent users (LRA<10) and the YFS arm, although this did not reach statistical significance. Of the 100 participants in the YFS+APP arm, 23 (23%) were interviewed. The risk assessment function is perceived as the most useful app feature. Further aesthetic adaptations and a more comprehensive rewards system were suggested by the interviewees.

Conclusions: Higher rates of PrEP adherence among frequent app users were observed; however, this was not statistically significant. A short app use duration of 3 months suggests that they may be useful in establishing habits in taking daily PrEP, but not long-term adherence. Further studies on the specific mechanisms of mobile phone apps that influence health behaviors are needed.

Trial Registration: ClinicalTrials.gov NCT03778892; https://clinicaltrials.gov/ct2/show/NCT03778892

(JMIR Mhealth Uhealth 2022;10(4):e25561) doi:10.2196/25561

KEYWORDS

mHealth; PrEP adherence; adolescents; men who have sex with men; transgender women; mobile phone

Introduction

Background

The emergence of advanced technologies and widespread use of smartphones provides opportunities for new HIV research and prevention strategies using mobile health (mHealth) within high-risk populations [1,2]. mHealth is the practice of medical and public health supported by mobile devices, such as mobile phones, patient monitoring devices, PDAs, and other wireless devices [3]. Data on user interaction with web-based intervention tools are readily available as part of such technologies and form useful surrogate markers of user engagement characteristics [4]. Despite the high potential for mHealth use in service delivery, there is a lack of data on how this is best delivered in low- and middle-income countries and in youth [1,2].

Men who have sex with men (MSM) and transgender women (TGW) are at high risk for HIV infection globally [5]. In Thailand, over 80% of new infections occur in this population [6], with young MSM (YMSM) and young transgender women (YTGW) at even greater risk than their older counterparts [7,8], as HIV incidence among YMSM and YTGW is 4-12 per 100 person-years [9-13], well above the 3 per 100 person-years World Health Organization incidence recommendation priority for offering HIV pre-exposure prophylaxis (PrEP) [14].

With the safety and effectiveness of PrEP established [14-18], the current challenge with PrEP is how best it can be implemented in key populations. The use of PrEP to prevent HIV has been recommended in the Thai national guidelines since 2014 [19], initially available as a fee-based PrEP the same year and also the following year through PrEP demonstration projects in 2015. The Princess PrEP Project, a key population-led PrEP service delivered to MSM and TGW by lay providers, was launched in 2016, contributing to 60% of all national PrEP uptakes [20,21]. *PrEP2start* was then launched by the Thai Ministry of Public Health in 2017, a nationwide scale-up initiative to increase accessibility to PrEP for key populations, including TGW and MSM [22]. Most recently, in 2019, PrEP became available under Thailand's Universal Health Coverage Scheme [21].

Given that PrEP efficacy is highly reliant on adherence [23], strategies to support adherence, encompassing those from biological, psychological, and social fronts, are key to its effectiveness at the public health level [24-27]. PrEP adherence in adolescents is known to be a challenge with adherence observed at 28% to 48% after 6 months of use, and it has been acknowledged that tailored approaches for this key population are needed to deliver effective prevention programs [25-28]. mHealth has been implemented in managing various health behaviors in adolescents, including sexual health promotion, disease prevention, antiretroviral adherence in HIV, and emotional health support, and outcomes have been promising [29,30]. A pilot mHealth study, iText, to support PrEP adherence motivation using weekly SMS text or email support messaging, found that it was acceptable, particularly among young participants, and demonstrated a 50% to 77% reduction in missed PrEP dosing with its use [31]. There is currently a lack of data evaluating eHealth service delivery strategies from low- to middle-income countries to support HIV prevention efforts [1,30,32]. Mobile phone health apps that use bidirectional interactions and are designed based on behavioral theories have been found to be more effective in influencing health behaviors than those that send unidirectional messages [2]. A frequently used theory in mobile app design is the information–motivation–behavioral (IMB) skills model that asserts that initiation and maintenance of health-promoting behaviors come about as a result of a combination of health-related IMB skills [2,33].

Objectives

Our aim in this study is to investigate participant engagement and the impact on PrEP adherence of the *Project Raincoat* mobile phone app, which was based on the IMB skills model developed to support PrEP adherence among YMSM and YTGW at risk of HIV acquisition, in the context of an adolescent-friendly clinic providing bidirectional web-based communications in Thailand. The primary results of this trial showed no difference in measured PrEP adherence between arms that received the app and those that did not; approximately 50% of all PrEP users achieved protective drug levels and no seroconversions were observed [33].

Methods

Overview and Ethical Considerations

This was a prospective randomized controlled trial of oral daily PrEP in youth at risk of HIV acquisition in Bangkok, Thailand [34]. Adolescents aged 15-19 years, assigned male gender at birth, and self-defining as MSM or TGW with HIV risk acquisition behaviors, defined as having >1 sex partner and inconsistent or no condom use in the preceding 6 months were included in this study. Participants could be new, current, or former PrEP users. They were randomized (1:1) to receive oral daily tenofovir-disoproxil fumarate/emtricitabine provided by either youth-friendly services (YFS) only or YFS plus the use of the *Raincoat* app (YFS+APP). This study was conducted in two different settings: (1) medical center–based facilities at the Thai Red Cross AIDS Research Center (TRCARC), the largest voluntary HIV testing center of Thailand, and the King Chulalongkorn Memorial Hospital, a major teaching hospital in Bangkok and (2) key population-led community-based drop-in centers operating as satellite sites of TRCARC, namely, the Rainbow Sky Association of Thailand (RSAT) and Service Workers In Group (SWING), both located in Bangkok, Thailand. Care at the 2 different medical center–based facilities was performed by the same team that moved between the 2 locations and tended to receive clients slightly more medically informed than those in the community-based drop-in centers. RSAT, owing to its proximity to universities and nearby student accommodation, received proportionally more cases than SWING; however, staff at both RSAT and SWING delivered similar care, with both teams receiving training from TRCARC and our study team on standard operating procedures.

The institutional review board approval was granted by the Faculty of Medicine, Chulalongkorn University (number 1091/2017), with a waiver for parental consent. This study was registered with ClinicalTrials.gov (NCT03778892).

Youth-Friendly Services

YFS included the following: (1) monthly engagement via either in-person clinic visits (months 1, 3, and 6) or telephone calls (months 2, 4, and 5) and (2) access to counselors and site staff outside scheduled visits through web-based messaging or telephone calls with responses provided within 24 hours. Clinic visits were available during weekdays and Saturday mornings to accommodate adolescent lifestyles. Motivational interviewing focused on HIV risk reduction and empowerment on using available HIV prevention methods was used by counselors during all interactions with clients [35,36].

Mobile Phone App Development and Use

The *Raincoat* mobile app was developed in conjunction with Focal Intelligence Co. Ltd. and designed using the IMB skills model guided by input from two adolescent focus group discussions (FGDs), one with 3 MSM and one with 3 TGW, all aged between 15 and 19 years [37,38]. Discussions were conducted using a semistructured interview guide with topics on desirable app functions, aesthetic preferences, potential barriers, and motivators for use. Key themes and subthemes were identified from content analysis, which then informed app design. The app prototype was tested with staff providing HIV prevention care to clients at all study sites (2 from each site, totaling 8 usability testers) whose feedback was used to inform the final app design.

The final app designed had three main features, which we refer to as the *3Rs*: *risk assessment*, *reminders*, and *rewards*. The former addresses the *information* component of the IMB model, and the latter two address the *motivation* component, with further explanation as follows: (1) *risk:* self-assessment of HIV acquisition risk with a once-weekly data input portal on the number of sex acts, sex partners, PrEP pills taken, and condom use, which was then used to calculate a feedback HIV risk level of low, medium, high, and very high; (2) *reminders:* in-built alarms for taking PrEP and HIV risk self-assessment using default set messages that were customizable; and (3) *rewards:* points were rewarded in real time for data input (maximum reward of 21 points per week) as well as to responding to staff follow-up calls (5 points each), attendance of clinic visits (10 points each), and negative anti-HIV test results (50 points each). Points rewarded were part of the intervention and were available to the YFS+APP arm only. Points were exchangeable for cash, with redemption being available at every 100 points if redeemed before the end of the study. Moreover, 100 points were exchangeable for 100 (US $3) at month 3 or month 6 clinic visits. A maximum of 719 points could be accumulated when using the app for 6 months.

The app was available only on the Android operating system. Android mobile phones were available for loans to those using other operating systems. Participants were free to log in to the app at any time. Staff provided guidance at their enrollment visit on how to load the app and introduced its functions. Participants required an email for initial registration and were asked to create their own usernames and passwords. Usernames and passwords were required for each log-in to ensure confidentiality. If participants had any issues using the app, they were free to contact the staff to resolve them. Customizable medication reminder messages could be created and changed at any time. The app was accessible offline for all features apart from the self-risk assessment feature, which could only be used with internet connectivity.

Mobile App Users—Exit Interviews

To solicit the attitudes of users toward the app, exit interviews were conducted using open-ended questions. Using convenience sampling, we interviewed a selection of participants who had (1) never logged in to the app, (2) used the app infrequently (defined as performance of log-in and risk assessment [LRA] of 1-9 times throughout the trial), and (3) been frequent users (defined as LRA≥10 times throughout the trial), with at least 10% of the total interviewees sampled from each group. Interview topics included facilitators and barriers to app use, valuable features of the app, reasons for nonuse, and general suggestions for improvement.

Study Procedures and Follow-up of PrEP Services

Follow-up clinic visits occurred at months 1, 3, and 6, and telephone contact was made at months 2, 4, and 5. Risk

behaviors and risk perception surveys were conducted for each monthly contact.

Substance use information was collected from self-reported surveys completed by participants privately via an electronic form consisting of alcohol, sildenafil citrate, and other recreational drugs (eg, amphetamines, methamphetamines, ketamine, poppers [volatile alkyl nitrates], and marijuana). HIV blood testing was performed at baseline and at months 1, 3, and 6.

Biological Measurement of PrEP Adherence

Dried blood spots (DBSs) were collected for quantification of tenofovir diphosphate (TFV-DP) concentrations at months 3 and 6 of follow-up. Whole blood samples were collected on Whatman Protein Saver 903 cards. DBSs were stored at −70 °C until analysis. TFV-DP was analyzed by liquid chromatography mass spectrometry. The TFV-DP calibration curve range was 200-10,000 fmol/3 mm punch [23,39] performed at the Program for HIV Prevention and Treatment Research Institute for Sustainable Development Pharmacology Laboratory at the Faculty of Associated Medical Sciences, Chiang Mai University. Details of the study methods have been previously published [34].

Statistical Analysis

Data on app use were downloaded for analysis after 6 months of follow-up. App paradata collected included first and last log-in dates, number of times logged in, number of LRAs, use of reminder messages, use of the medication alarm function, and total points accumulated.

The number of LRAs was used as an independent variable to assess user engagement with the app. TFV-DP DBS concentrations were used to evaluate PrEP adherence, with TFV-DP levels ≥700 fmol/punch considered good adherence (equivalent to ≥4 tablets/week) [23].

Continuous variables were presented as means with SDs or medians with IQRs, and categorical variables with absolute numbers and percentages. The chi-square test, Z test, Fisher exact test, odds ratios, and 95% CIs were used for group comparisons and associations as appropriate. Stata/SE (version 13.0; StataCorp) was used for quantitative data analyses.

Results

Overview

Details of baseline characteristics and comparisons between the control arm (YFS) and the intervention arm (YFS+APP) have previously been published [34]. In brief, between March 2018 and June 2019, 200 HIV-negative participants were enrolled, 100 received YFS and 100 YFS+APP. The median age at enrollment was 18 (IQR 17-19) years, with 73.5% (147/200) self-defining as YMSM and 26.5% (53/200) as YTGW. Participants were enrolled and followed-up at medical center–based facilities (135/200, 67.5%) or community-based drop-in centers (65/200, 32.5%). Baseline characteristics between the YFS and YFS+APP arms were similar except for self-reported substance use in the preceding 3 months, which was higher in the YFS+APP arm, 18% (18/100), than in the YFS arm, 8% (8/100; *P*=.04) [34]. Of the 143 adolescents reporting sexual activity in the past month, 94 (65.7%) reported inconsistent condom use. Of the 200 participants, 187 (93.5%) rated themselves as having a low risk of HIV acquisition. There were no significant differences in PrEP adherence between the YFS and YFS+APP arms. PrEP adherence was 51% (40/79) in the YFS arm and 54% (44/81) in the YFS+APP arm (*P*=.64) at month 3 and was 44% (30/68) in the YFS arm and 49% (36/73) in the YFS+APP arm (*P*=.54) at month 6, further details of which have previously been published [34].

Risk Assessments

Of the 100 participants randomized to receive YFS+APP, 87 (87%) used LRA at least once during the follow-up period. Of these 87 app users, 55 (63%) used their own phones, and the remaining used loaned Android operating system phones. Of the 13 participants who never logged into the app, 10 (77%) used their own phones and 3 (23%) used loaned phones. Among the app users, the median (IQR) duration between the first and last LRA was 3.5 (1.6-5.6) months, with a median LRA frequency of 6 (IQR 2-10). There was no difference in median LRA frequency between participants who used their phones and those who used loaned phones (*P*=.21). The percentage of participants who used LRA declined over time, that is, 77%, 52%, and 28% at 2, 6, and 12 weeks after enrollment, respectively (Figure 1).

Of the 87 app users, 65 (76%) were infrequent users (LRA<10). Associations of LRA frequency and baseline characteristics among app users were assessed, and only gender identity was significantly associated with LRA, with YMSM 9.3 times (95% CI 1.2-74.3; *P*=.04) more likely to be frequent users than YTGW, but with a very wide CI (Table 1).

Associations between LRA frequency and PrEP adherence determined by TFV-DP DBS levels at months 3 and 6 are shown in Table 2.

Figure 1. Percentage of participants who logged in and performed risk assessments by week after enrollment (N=87).

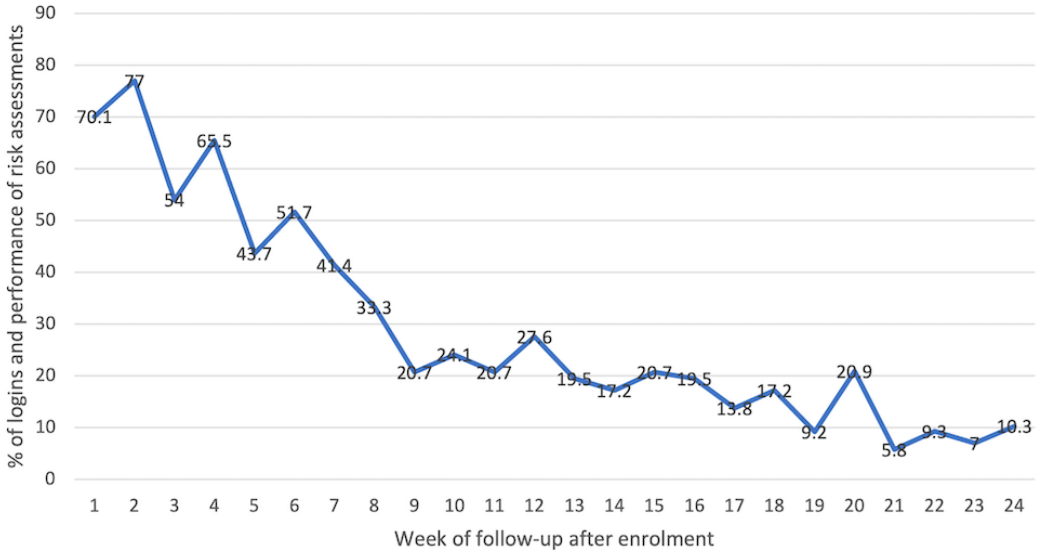


Table 1. Baseline characteristics of participants in the YFS+APP[a] arm who logged in to the mobile phone app and performed risk assessment (LRA[b]) at least once during the trial follow-up period and associations with LRA.

Demographic characteristics	Total (n=87), n (%)	LRA≥10 times, n (% by characteristics)	OR[c] (95% CI)	*P* value
Gender identity				
MSM[d]	66 (75)	21 (31.8)	9.3 (1.2-74.3)	.04
TGW[e]	21 (24)	1 (4.8)	1	N/A[f]
Age at enrollment (years)				
15-17	29 (33)	6 (20.7)	1	N/A
18-19	58 (66)	16 (27.6)	1.5 (0.5-4.2)	.49
Number of sex partner or partners in the last month				
0	25 (28)	8 (32)	1	N/A
1	38 (43)	12 (31.5)	1.0 (0.3-2.9)	.97
≥2	24 (27)	2 (8.3)	0.2 (0.0-1.0)	.05
Condom use in the past month among participants reporting sexual activity (n=62)				
100	18 (29)	4 (22.2)	1	N/A
<100	44 (71)	10 (22.7)	1.0 (0.3-3.8)	.97
Recreational substance use in the past 3 months (n=62)				
No	70 (80)	19 (27.1)	1	N/A
Yes	17 (19)	3 (17.6)	0.6 (0.2-2.2)	.42
Self-perceived risk (n=62)				
Low or moderate risk	52 (59)	16 (30.8)	1	N/A
High risk	35 (40)	6 (17.1)	0.5 (0.2-1.34)	.16

[a]YFS+APP: youth-focused strategies plus mobile phone app.

[b]LRA: log-in and risk assessment.

[c]OR: odds ratio.

[d]MSM: men who have sex with men.

[e]TGW: transgender women.

[f]N/A: not applicable (reference).

Table 2. Associations of Raincoat mobile phone app use and HIV PrEP[a] adherence (tenofovir diphosphate level ≥700 fmol/punch) at months 3 and 6.

App users	PrEP adherent, n (%; 95% CI)	Unadjusted		Adjusted[b]	
		Odds ratio (95% CI)	*P* value	Odds ratio[c] (95% CI)	*P* value
At 3 months					
YFS[d] only (n=79)	40 (50; 39.6-61.6)	1.0	N/A[e]	N/A	N/A
YFS and infrequent app use (n=53)	29 (55; 41.3-68.1)	1.8 (0.6-2.4)	.64	1.1 (0.5-2.4)	.84
YFS and frequent app use (n=21)	14 (67; 46.5-86.8)	2.0 (0.7-5.4)	.19	1.8 (0.6-5.4)	.59
At 6 months					
YFS only (n=68)	30 (44; 32.3-55.9)	1.0	N/A	N/A	N/A
YFS and infrequent app use (n=47)	21 (45; 30.5-58.9)	1.0 (0.5-2.2)	.95	0.8 (0.4-1.9)	.64
YFS and frequent app use (n=22)	13 (59; 38.5-79.6)	1.8 (0.7-4.8)	.22	1.8 (0.6-5.3)	.27

[a]PrEP: pre-exposure prophylaxis.

[b]Adjusted by gender identity (men who have sex with men vs transgender women), age at enrollment (15-17 vs 18-19) (years), number of sex in the past month at enrollment interviewed (0 vs 1 and ≥2), and self-perceived HIV acquisition risk at enrollment.

[c]Unadjusted odds ratio (95% CI) of adjusted variables in footnote[a], as previously published [34].

[d]YFS: youth-friendly services.

[e]N/A: not applicable (reference).

The overall percentage of PrEP adherence was higher in frequent app users than in infrequent app users and the YFS arm, but this did not reach statistical significance. At month 6, the proportion of participants who achieved TFV-DP≥700 fmol/punch was 59% (13/22), 44% (30/68) among frequent app users, 45% (21/47) for infrequent users, and 44% (30/68) in the YFS arm (*P*=.47).

Exit Interview Findings

A total of 23 adolescents in the YFS+APP arm participated in exit interviews, of which 18 (78%) were YMSM and 20 (87%) were active PrEP users at the time of interview. Of all 23 participants interviewed, 3 (13%) had never logged in to the app, 17 (74%) were infrequent users, and 3 (13%) were frequent users. Interviews were conducted at a median (IQR) of 8.5 (5.0-12.4) months after completion of the study.

Reasons for app nonuse were categorized into those that were app-specific and those that were lifestyle-related. App-specific factors included the inconvenience of using a borrowed phone, requiring email verification during the registration process, finding app aesthetics and functions unengaging, and its lack of complete offline functionality. Lifestyle factors included changing phones, not loading the app onto their new phone, and being too busy with work and studies to use the app.

Regarding the risk assessment app feature, participants liked this feature the most because of the colorful graphics used in feeding back risk levels, because of not requiring the answering of too many questions, and also because it helped them to be more aware of their HIV risk. All 20 app user interviewees had used risk assessment features and reported that they performed risk assessments in the beginning and less later on for multiple reasons, including not being sexually active during the time, shift work, being busy with studies, forgetting to take their phone with them, or just forgetting to do the assessment.

Interviewees emphasized that for an app to be desirable for adolescents, it would need to be easy to use, feature attractive aesthetic features such as customizable color schemes and customizable avatars, and also provide the opportunity to interact with others. Privacy was also another major concern, and passcode protection and discreet branding of the app were also important. Although many participants expressed a desire for the app to have more information owing to its convenience and credibility, some felt that searching for information themselves on the web was easier and also provided the most up-to-date information. The app also needed to be competitive with other available apps, both health-related and nonhealth-related, in the market.

Reminders for Taking PrEP and Visit Reminders

Of the 87 app users, 45 (52%) used PrEP reminders and 34 (39%) clinic appointment reminders. Of the 45 participants who used PrEP reminders, 8 (18%) customized their reminder messages. Proportions of PrEP adherence observed between those who used and did not use the PrEP reminder function did not differ at month 3 (50% vs 68%; *P*=.13) or month 6 (50% vs 53%; *P*=.82).

Of the 20 app user exit interviewees, 15 (75%) reported using the PrEP reminder feature, and of the 15 reminder users, 8 (53%) liked the function because it helped them take PrEP on time. Moreover, 1 participant viewed this function as the most important feature of the app. Some interviewees who stopped using this function said they switched to using their phone alarm as it was more convenient, particularly in those using a borrowed study phone, as sometimes the app alarm did not go off as set.

Rewards and Redemption

Of the 87 app users, 51 (59%) accumulated sufficient points for cash redemption (range 100-504 points). Among the users, 49% (43/87) earned 100 to 199 points, 33% (29/87) earned 200 to

299 points, 16% (14/87) earned 300 to 399 points, and 2% (2/87) earned ≥400 points.

For redeeming prizes, exit interviewees, with 65% (13/20) of them having had participated in cash redemption, wanted the entire process to be automated (although accumulated points were visible in the app, redemption had to be done manually at clinic visits by study staff). The cash rewards used in this app were liked by 25% (5/20) of the app users, but many felt there was a need to increase variation in rewards with additional gimmicks, such as exchange of points for other prizes such as movie tickets and food vouchers. More importantly, to motivate the use of the app, the timing of reward redemption they felt should be earlier than at the 3- and 6-month clinic visits in this study.

Engagement With Site Staff Outside Monthly Clinic or Phone Call Visits

There were 578 interactions between study staff and 140 participants, with a median (IQR) of 4 (3-5) times per person. Engagement with staff was similar between the YFS (67/100, 67%) and YFS+APP (73/100, 73%) arms. Of the 578 interactions, 299 (51.7%) were initiated by clients. The median (IQR) time to first contact after enrollment was 6 (3-13) weeks. The topics of interactions included making appointments (128/299, 42.8%), relationships and personal issues (70/299, 23.4%), PrEP inquiries (54/299, 18.1%), HIV risk (29/299, 9.7%), and sexually transmitted disease diagnosis and management (18/299, 6%; Table 3).

Regarding what service aspects would keep them coming back, many felt this ultimately came down to staff being friendly, open, approachable, reliable, understanding, and holistic in their care.

Table 3. Topics of inquiry initiated by participants to site staff during 6 months of follow-up.

Contact topics	Participants initiating contact (n=140[a]), n (%)[b]	Contacts (n=299), n (%)	YFS[c] (n=149), n (%)	YFS+APP[d] (n=150), n (%)	*P* value
Clinic appointments	90 (64.3)	128 (42.8)	64 (43)	64 (42.7)	.96
Relationships and personal issues	62 (37.1)	70 (23.4)	28 (18.8)	42 (28)	.06
HIV PrEP[e]	44 (31.4)	54 (18.1)	34 (22.8)	20 (13.3)	.03
HIV risk acquisition	26 (18.6)	29 (9.7)	14 (9.4)	15 (10)	.86
Sexually transmitted diseases	11 (7.8)	18 (6)	9 (6)	9 (6)	.99

[a]Total number of participants initiating contact.

[b]Some participants inquired more than once on >1 topic.

[c]YFS: youth-friendly services.

[d]YFS+APP: youth-friendly services plus mobile phone app.

[e]PrEP: pre-exposure prophylaxis.

Discussion

Principal Findings

This randomized controlled study examined how a smartphone PrEP adherence support app used among adolescents in Thailand affected PrEP adherence and observed higher proportions of PrEP adherence among frequent app users than among infrequent users and the YFS arm at both months 3 and 6, although this did not reach statistical significance. To our knowledge, this study is one of the few studies that address key populations where a dearth of information exists in both young adolescents and low- and middle-income settings to inform future policy and practices in such settings [2,30,40]. Although some degree of success in using telehealth at public health levels in Thailand has been seen in some studies in Thailand, including the use of mobile phone SMS text messaging to support PrEP adherence [41], telephone-based drug adherence support for tuberculosis [42], web-based recruitment and linkage to home-based HIV testing [43], the use of mHealth technologies in HIV prevention efforts in Thailand and the Asia Pacific region is in its early stages, and few publications beyond preimplementation studies exist regarding mobile phone app implementation and effectiveness in risk behavior modification. The app produced and used in this trial was based on a behavioral change theory, IMB, where *information* provided on risk with self-assessments and *motivation* with reminders with medication alarms and point rewards.

Risk Assessment

Of the 87 participants, 22 (25%) were classified as frequent users, reflecting that app use was short-term in most users, which is consistent with a previous observation that 33% of users stop using their device in <6 months and 39% of commercial health apps were used ≤10 times before use was discontinued [44]. The lack of additional PrEP adherence benefit seen with mobile phone app use in this trial may have been due to *dilution* by the large proportion of infrequent app users and nonapp users in those randomized to use it. Given that it is known that, globally, mobile app retention, defined as the use of an app in the preceding 7 days, is only approximately 15%, the challenge of measuring real-world mobile phone app effectiveness could be addressed with larger trials to establish whether this short duration of use is sufficient to elicit behavioral change or whether strategies are needed to encourage longer use to enable behavioral change [45]. It is also important to consider whether app use is able to elicit behavioral change, specifically by what mechanism, which could be done by isolation and testing of

individual components to allow future innovations to build upon in developing effective strategies, currently a research gap in mHealth technologies [32,44,46]. Improvement of aesthetics and ensuring features were customizable, having more options for cash reward exchanges, having more social interaction, and more health information were also major themes that arose in the exit interviews as areas in which users felt the app could be improved to increase user engagement. This is consistent with a previous review of mobile phone app use in HIV prevention that app inclusiveness and interactivity are more likely to attract and retain users as well as elicit behavioral change [45].

Individuals may have felt no need to continue performing self-risk assessments if they intended to continue the same preventive measures. Those who continued with LRA may have been those who had fluctuations in sexual HIV risk behaviors, the *worried well*, or point reward collectors. Many exit interviewees said that the action of assessment was a reminder to consider self-risk, which in turn influenced planned health behaviors. We feel this self-acknowledgment of one's true risk level was supported by the app, which enabled privacy and discrete assessment and bypassed the need for human interaction and potential judgmental attitudes faced in making this assessment. In addition, users were able to try inputting data relevant to different scenarios, as it was possible to repeat assessments in planning upcoming sexual encounters. We think that the large drop-off of users toward the end of the trial is reflective of findings from previous studies that apps are mainly effective in those who are already motivated but are less effective at supporting user capabilities and maintaining behavioral changes [46].

On the basis of the exit interview findings, the *Raincoat* app was found to be easy to use. App design allowing ease of utility clearly influenced how likely participants would use the app. This observation is in keeping with previous studies that found simple interfaces may reduce the time participants need to spend on the app and therefore improve retention [30].

Reminders

Over half of the app users used their PrEP reminder function. Our exit interviews found that many users felt that the function was very useful for supporting PrEP adherence. Tailored, short pieces of information on self-risk fed back at weekly intervals used in this trial has been reported to be the preferred mode of messaging in other studies compared with general health information [46]. However, in this study, no association was observed between medication alarm use and PrEP adherence, suggesting that this feature alone may not support PrEP adherence for all users. Furthermore, given that only 18% (8/45) of participants customized their reminder message function, it could be argued that this is a nonessential feature for future versions. All those interviewed said that the alarm function did not always work in this trial was a barrier to their engagement with it. This study highlighted the implementation of mHealth technologies, particularly in low- and middle-income settings. Plans and budget allocations must be made for the technical maintenance of their functionality to ensure engagement and maximum benefit of use over time. It should also be noted that some participants switched to using their own phone alarm for medication reminders, suggesting that the alarm function may not be an essential in-app feature for adolescents, particularly in resource-limited settings.

Rewards and Redemption

In this study, we observed that cash rewards may not be sufficient on their own to influence health behaviors, which is consistent with a previous review that suggested, at best, conditional economic incentives can be a valuable part of but cannot replace other HIV prevention approaches and, importantly, need to be adapted to specific contexts and populations for maximal success [47]. This would need to be carefully experimentally isolated to explore this possibility further [48].

Engagement With Site Staff

Engagement in this study at monthly scheduled visits or phone calls plus optional additional staff consultation producing a 73% retention rate may have reduced the impact of receiving the app intervention seen, given that it is known that 2-way interaction between providers and clients is known to improve medication adherence [2]. Appointment-related queries made up nearly half of all contact topics in both study arms and accounted for 64.3% (90/140) of all participants who initiated contact. This strongly suggests that an automated system in the app to book and confirm appointments would improve convenience for both service users and providers. Relationships and personal issues and PrEP-related queries also made up approximately one-fifth each of queries to staff, suggesting that specific in-app information provision and staff training on these issues may benefit service users. These data highlight the unique considerations to be made in adolescent and young adult HIV prevention service provision to address developmental needs to support their decision-making, establishment of identity, interpersonal challenges, and mental health issues commonly seen at this stage of life [27].

Types of Clinic Settings or Services

Although the intervention of interest was a mobile phone app, it must be acknowledged that the contexts in which this study occurred were medical center–based facilities with HIV expertise or community-based organizations (CBOs), both of which are unique to general PrEP provision services available in Thailand. All the locations used in this study have study staff with training on the same basis to deliver the intervention and a focus of work on HIV prevention and client-centered care, with medical center–based facilities having the strength of delivering the intervention directly and CBOs having the advantage of key population-led care, which, as has been seen in other studies, supported PrEP adherence and retention in services well [20,49]. It is possible that such differences could have acted as effect modifiers in the interventions studied. General PrEP services in Thailand are delivered at medical center–based facilities with a larger variety of health issues and have a larger volume of service users, making holistic health care much more challenging to deliver. Medical centers and CBOs may draw different types of populations, and with randomization, the characteristics are balanced between the 2

arms. The heterogeneity of the sample makes the generalizability of the study results to broader populations.

It is also recognized that mHealth has greatly facilitated the maintenance of some health care delivery services in the last year during the COVID-19 pandemic [50-53]. Evidence suggests that HIV at-risk populations globally, including YMSM and YTGW, have faced considerable health care access disruptions and urgent development and implementation of telehealth as part of efforts to minimize such disruptions, in addition to other policy level initiatives, will minimize associated care disruption-related HIV morbidity and mortality [54,55]. The largest PrEP provider of Thailand, the *Princess PrEP Program*, has recognized since its initiation in 2016 that service users require adherence and retention support, particularly in TGW, which could be supported using mHealth technologies [21].

Limitations

Although the original design of this study was fully powered to compare those who received YFS versus YFS+APP, as a significant proportion of study participants infrequently used the app, it became necessary for the final analysis to stratify outcome groups into frequent users (LRA10) and infrequent users (LRA<10). The actual analysis was therefore not fully powered; thus, results from this study should be viewed as preliminary as part of a pilot randomized control trial. Another limitation of this study was that the app used for the intervention in this study was based on just 2 FGDs, which may have provided an inadequate range of views in the design of the app. In addition, the protocol of this study was designed to describe the *natural history* of app use without any formal encouragement from the study team. App paradata were therefore collated at the end of study for this analysis. However, another possible policy that could have been taken was to continually look at app use and discuss app use obstacles at each contact to rapidly assess and address functionality problems and also to collect design issues *real time* when these issues are *fresh* in the minds of the users and more likely to be recalled more clearly. In addition, given that the efficacy of mobile technologies is also linked to literacy and the performance of this study in a large city, findings from it may not apply to a more rural area where access to mobile technologies and literacy may be more challenging and less benefit gained from such a form of service provision enhancement, a limitation of mHealth seen previously [44]. Only a total of 6 interviewees participated in FGDs that informed the design of this app, which may have affected the generalizability of the preference representation drawn from this sample of adolescents. The sampling frame used for exit interviews in this study was also biased toward those who were classified as frequent users of the app, which may have led to findings being more biased toward favoring app functions. Given no differences were seen between arms, interviews with control arm participants could have been done to provide further information on possible reasons for this. Interviews in this study were conducted for quite a long time after study completion in most cases, which may have led to limited data quality owing to implications on participant recall. The small sample size of this study also limits the generalizability of the conclusions drawn from this study, and it is possible that the intervention was not efficacious or the sample size was too small to see any effect present.

Conclusions

Using a target population-informed design for a mobile phone app supporting PrEP adherence in an adolescent PrEP implementation study, we observed higher rates of PrEP adherence among those who frequently used the app, although the difference was not statistically significant. App use dropped by 50% by week 6, suggesting its possible usefulness in establishing habits but not for long-term adherence maintenance. Further studies on specific mechanisms in which mobile apps are effective in influencing health behaviors are needed to inform future mHealth scale-up operations, particularly in low- and middle-income settings with limited staff service capacity. mHealth development and implementation will ultimately benefit health access and delivery for HIV prevention services beyond its current apps in the current COVID-19 pandemic.

Acknowledgments

The authors would like to acknowledge members of the Center of Excellence for Pediatric Infectious Diseases and Vaccines—Institute of HIV Research and Innovation Adolescent Study Team: Praphan Phanuphak, Chitsanu Pancharoen, Somsong Teeratakulpisarn, Reshmie Ramautarsing, Rena Janamnuaysook, Kritima Samitpol, Jiratchaya Kongkapan, Artsanee Chancham, Pravit Mingkwanrungruang, Narukjaporn Thammajaruk, Nuttawut Teachatanawat, Krittaporn Termvanich, Tippawan Pankam, Thantip Sungsing, Sorawit Amatavete, Chotika Prabjantuek, Phubet Panpet, Jureeporn Jantarapakde, Tuangtip Theerawit, Watsamon Jantarabenjakul, Suvaporn Anugulruengkitt, Nattapong Jitrunruengnit, Noppadol Wacharachaisurapol, Pintip Suchartlikitwong, Pakpoom Janewongwirot, Krittaporn Pornpaisalsakul, Juthamanee Moonwong, Rachaneekorn Nadsasarn, Patchareeyawan Srimuan, Siwanart Thammasala, Sasiwimol Ubolyam, Patcharin Eamyoung, Jiratchaya Sophonphan, Stephen Kerr, Taninee Petwijit, Chansuda Bongsebandhu-phubakdi, Buajoom Raksakul, Pathranis Meekrua, Orawan Fungfoosri, Prapaipan Plodgratoke, Pakitta Kidsumret, Kamolchanok Chumyen, Sahakun Chintanakarn, Sirinthon Yenjit, Chanin Suksom, Tanunjit Rugphan, Somsri Tantipaibulvut, Prakaipech Kaw-In, Penprapa Chanachon, Chanjiraporn Pondet.

The authors also thank Pathomchai Amornratpaijit from the Center of Excellence for Pediatric Infectious Diseases and Vaccines for his support in finance and administrative management; Focal Intelligence Co. Ltd team members Rachapong Pornwiriyangkura, Mathawee Tusjunt, Patcharamon Siriphongpiphat, Rawisara Loekiatthamrong, and Sitthiphong Achavan for their support in designing and producing *Project Raincoat*; Yardpiroon Tawon for her hard work in supporting tenofovir diphosphate assay analyses; Dr Sara LeGrand and Shashika Bandara of Duke University for their support in conducting the Chulalongkorn University

mHealth training and support in writing this manuscript; and Dr. Arunrat Tangmunkongvorakul of Chiang Mai University for her support in conducting qualitative study capacity training for the study staff.

CIPHER (Collaborative Initiative for Pediatric HIV Education and Research; grant 2017/472-SON) provided core funding for this study. The medication and medical care provided through the Princess PrEP Program was supported by the US President's Emergency Plan for AIDS Relief and the US Agency for International Development through the Linkages Across the Continuum of HIV Services for Key Populations Cooperative Agreement (grant AID-OAA-A-14-0045) managed by FHI 360. The Ratchadapisek Sompoch Endowment Fund (2019-2020) under Telehealth Cluster, Chulalongkorn University, provided supplemental mHealth Research Funding Support and also funding for mHealth staff capacity building, The Chulalongkorn University C2F Postdoctoral Fellowship Fund supported SK, and The Chulalongkorn University Ratchadapisek Sompoch Postdoctoral Fellowship Fund (2019-2020) supported WNS.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT eHEALTH Checklist (V 1.6.1).
[PDF File (Adobe PDF File), 175 KB - mhealth_v10i4e25561_app1.pdf]

References

1. Nguyen LH, Tran BX, Rocha LE, Nguyen HL, Yang C, Latkin CA, et al. A systematic review of eHealth interventions addressing HIV/STI prevention among men who have sex with men. AIDS Behav 2019 Sep;23(9):2253-2272 [FREE Full text] [doi: 10.1007/s10461-019-02626-1] [Medline: 31401741]
2. Henny KD, Wilkes AL, McDonald CM, Denson DJ, Neumann MS. A rapid review of eHealth interventions addressing the continuum of HIV care (2007-2017). AIDS Behav 2018 Jan;22(1):43-63 [FREE Full text] [doi: 10.1007/s10461-017-1923-2] [Medline: 28983684]
3. WHO Global Observatory for eHealth. mHealth: New Horizons for Health Through Mobile Technologies: Second Global Survey on eHealth. Geneva: World Health Organization; 2011.
4. Couper MP, Alexander GL, Zhang N, Little RJ, Maddy N, Nowak MA, et al. Engagement and retention: measuring breadth and depth of participant use of an online intervention. J Med Internet Res 2010;12(4):e52 [FREE Full text] [doi: 10.2196/jmir.1430] [Medline: 21087922]
5. Topics: key populations. UNAIDS. 2020. URL: https://www.unaids.org/en/topic/key-populations [accessed 2022-01-07]
6. van Griensven F, Guadamuz TE, de Lind van Wijngaarden JW, Phanuphak N, Solomon SS, Lo Y. Challenges and emerging opportunities for the HIV prevention, treatment and care cascade in men who have sex with men in Asia Pacific. Sex Transm Infect 2017 Aug;93(5):356-362. [doi: 10.1136/sextrans-2016-052669] [Medline: 28729520]
7. van Griensven F, Thienkrua W, McNicholl J, Wimonsate W, Chaikummao S, Chonwattana W, et al. Evidence of an explosive epidemic of HIV infection in a cohort of men who have sex with men in Thailand. AIDS 2013 Mar 13;27(5):825-832. [doi: 10.1097/QAD.0b013e32835c546e] [Medline: 23169330]
8. Idele P, Gillespie A, Porth T, Suzuki C, Mahy M, Kasedde S, et al. Epidemiology of HIV and AIDS among adolescents: current status, inequities, and data gaps. J Acquir Immune Defic Syndr 2014 Jul 01;66 Suppl 2:144-153. [doi: 10.1097/QAI.0000000000000176] [Medline: 24918590]
9. Thienkrua W, van Griensven F, Mock PA, Dunne EF, Raengsakulrach B, Wimonsate W, et al. Young men who have sex with men at high risk for HIV, Bangkok MSM cohort study, Thailand 2006-2014. AIDS Behav 2018 Jul 14;22(7):2137-2146 [FREE Full text] [doi: 10.1007/s10461-017-1963-7] [Medline: 29138981]
10. van Griensven F, Varangrat A, Wimonsate W, Tanpradech S, Kladsawad K, Chemnasiri T, et al. Trends in HIV prevalence, estimated HIV incidence, and risk behavior among men who have sex with men in Bangkok, Thailand, 2003–2007. J Acquir Immune Defic Syndr 2010 Feb;53(2):234-239. [doi: 10.1097/QAI.0b013e3181c2fc86] [Medline: 19901844]
11. HIV Surveillance Report, 2018 (Updated). Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/hiv/library/reports/hiv-surveillance.html [accessed 2022-01-17]
12. Wasantioopapokakorn M, Manopaiboon C, Phoorisri T, Sukkul A, Lertpiriyasuwat C, Ongwandee S, et al. Implementation and assessment of a model to increase HIV testing among men who have sex with men and transgender women in Thailand, 2011-2016. AIDS Care 2018 Oct;30(10):1239-1245 [FREE Full text] [doi: 10.1080/09540121.2018.1492697] [Medline: 29950108]
13. Colby D, Nguyen NA, Le B, Toan T, Thien DD, Huyen HT, et al. HIV and Syphilis prevalence among transgender women in Ho Chi Minh City, Vietnam. AIDS Behav 2016 Dec;20(Suppl 3):379-385. [doi: 10.1007/s10461-016-1485-8] [Medline: 27435074]

14. Policy brief: pre-exposure prophylaxis (PrEP): WHO expands recommendation on oral pre-exposure prophylaxis of HIV infection (PrEP). Document number WHO/HIV/2015.48. World Health Organization. 2015. URL: https://apps.who.int/iris/handle/10665/197906 [accessed 2022-01-20]
15. Grant RM, Lama JR, Anderson PL, McMahan V, Liu AY, Vargas L, et al. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. N Engl J Med 2010 Dec 30;363(27):2587-2599 [FREE Full text] [doi: 10.1056/NEJMoa1011205] [Medline: 21091279]
16. Hodges-Mameletzis I, Dalal S, Msimanga-Radebe B, Rodolph M, Baggaley R. Going global: the adoption of the World Health Organization's enabling recommendation on oral pre-exposure prophylaxis for HIV. Sex Health 2018;15(6):489. [doi: 10.1071/sh18125]
17. McCormack S, Dunn DT, Desai M, Dolling DI, Gafos M, Gilson R, et al. Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from the pilot phase of a pragmatic open-label randomised trial. Lancet 2016 Jan 2;387(10013):53-60 [FREE Full text] [doi: 10.1016/S0140-6736(15)00056-2] [Medline: 26364263]
18. Molina J, Capitant C, Spire B, Pialoux G, Cotte L, Charreau I, et al. On-demand preexposure prophylaxis in men at high risk for HIV-1 infection. N Engl J Med 2015 Dec 3;373(23):2237-2246. [doi: 10.1056/NEJMoa1506273] [Medline: 26624850]
19. Manosuthi W, Ongwandee S, Bhakeecheep S, Leechawengwongs M, Ruxrungtham K, Phanuphak P, Adults and Adolescents Committee of the Thai National HIV Guidelines Working Group. Guidelines for antiretroviral therapy in HIV-1 infected adults and adolescents 2014, Thailand. AIDS Res Ther 2015;12:12 [FREE Full text] [doi: 10.1186/s12981-015-0053-z] [Medline: 25908935]
20. Phanuphak N, Sungsing T, Jantarapakde J, Pengnonyang S, Trachunthong D, Mingkwanrungruang P, et al. Princess PrEP program: the first key population-led model to deliver pre-exposure prophylaxis to key populations by key populations in Thailand. Sex Health 2018 Nov;15(6):542-555. [doi: 10.1071/SH18065] [Medline: 30249317]
21. Ramautarsing RA, Meksena R, Sungsing T, Chinbunchorn T, Sangprasert T, Fungfoosri O, et al. Evaluation of a pre-exposure prophylaxis programme for men who have sex with men and transgender women in Thailand: learning through the HIV prevention cascade lens. J Int AIDS Soc 2020 Jun;23 Suppl 3:e25540 [FREE Full text] [doi: 10.1002/jia2.25540] [Medline: 32602660]
22. Suraratdecha C, Stuart RM, Manopaiboon C, Green D, Lertpiriyasuwat C, Wilson DP, et al. Cost and cost-effectiveness analysis of pre-exposure prophylaxis among men who have sex with men in two hospitals in Thailand. J Int AIDS Soc 2018 Jul 22;21 Suppl 5:e25129 [FREE Full text] [doi: 10.1002/jia2.25129] [Medline: 30033559]
23. Grant RM, Anderson PL, McMahan V, Liu A, Amico KR, Mehrotra M, et al. Uptake of pre-exposure prophylaxis, sexual practices, and HIV incidence in men and transgender women who have sex with men: a cohort study. Lancet Infect Dis 2014 Sep;14(9):820-829. [doi: 10.1016/S1473-3099(14)70847-3] [Medline: 25065857]
24. Mullins TL, Lehmann CE. Oral Pre-Exposure Prophylaxis (PrEP) for HIV prevention in adolescents and young adults. Curr Pediatr Rep 2018 Jun;6(2):114-122 [FREE Full text] [doi: 10.1007/s40124-018-0163-x] [Medline: 30345163]
25. Hosek SG, Landovitz RJ, Kapogiannis B, Siberry GK, Rudy B, Rutledge B, et al. Safety and feasibility of antiretroviral preexposure prophylaxis for adolescent men who have sex with men aged 15 to 17 years in the United States. JAMA Pediatr 2017 Nov 01;171(11):1063-1071. [doi: 10.1001/jamapediatrics.2017.2007] [Medline: 28873128]
26. Hosek SG, Rudy B, Landovitz R, Kapogiannis B, Siberry G, Rutledge B, Adolescent Trials Network (ATN) for HIVAIDS Interventions. An HIV preexposure prophylaxis demonstration project and safety study for young MSM. J Acquir Immune Defic Syndr 2017 Jan 01;74(1):21-29. [doi: 10.1097/QAI.0000000000001179] [Medline: 27632233]
27. Hosek S, Pettifor A. HIV prevention interventions for adolescents. Curr HIV/AIDS Rep 2019 Feb;16(1):120-128 [FREE Full text] [doi: 10.1007/s11904-019-00431-y] [Medline: 30707399]
28. Seekaew P, Nguyen E, Sungsing T, Jantarapakde J, Pengnonyang S, Trachunthong D, et al. Correlates of nonadherence to key population-led HIV pre-exposure prophylaxis services among Thai men who have sex with men and transgender women. BMC Public Health 2019 Mar 21;19(1):328 [FREE Full text] [doi: 10.1186/s12889-019-6645-0] [Medline: 30898095]
29. Badawy SM, Barrera L, Sinno MG, Kaviany S, O'Dwyer LC, Kuhns LM. Text messaging and mobile phone apps as interventions to improve adherence in adolescents with chronic health conditions: a systematic review. JMIR Mhealth Uhealth 2017 May 15;5(5):e66 [FREE Full text] [doi: 10.2196/mhealth.7798] [Medline: 28506955]
30. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
31. Fuchs JD, Stojanovski K, Vittinghoff E, McMahan VM, Hosek SG, Amico KR, et al. A mobile health strategy to support adherence to antiretroviral preexposure prophylaxis. AIDS Patient Care STDS 2018 Mar;32(3):104-111 [FREE Full text] [doi: 10.1089/apc.2017.0255] [Medline: 29565183]
32. Maloney KM, Bratcher A, Wilkerson R, Sullivan PS. Electronic and other new media technology interventions for HIV care and prevention: a systematic review. J Int AIDS Soc 2020 Jan;23(1):e25439 [FREE Full text] [doi: 10.1002/jia2.25439] [Medline: 31909896]

33. Fisher WA, Fisher JD, Harman J. The information-motivation-behavioral skills model: a general social psychological approach to understanding and promoting health behavior. In: Social Psychological Foundations of Health and Illness. Hoboken, New Jersey, United States: John Wiley & Sons; 2003:82-106.
34. Songtaweesin WN, Kawichai S, Phanuphak N, Cressey TR, Wongharn P, Saisaengjan C, CE-PID - TRC Adolescent Study Team. Youth-friendly services and a mobile phone application to promote adherence to pre-exposure prophylaxis among adolescent men who have sex with men and transgender women at-risk for HIV in Thailand: a randomized control trial. J Int AIDS Soc 2020 Sep;23 Suppl 5:e25564 [FREE Full text] [doi: 10.1002/jia2.25564] [Medline: 32869511]
35. Miller WR, Rollnick S. Meeting in the middle: motivational interviewing and self-determination theory. Int J Behav Nutr Phys Act 2012;9:25 [FREE Full text] [doi: 10.1186/1479-5868-9-25] [Medline: 22385872]
36. Palacio A, Garay D, Langer B, Taylor J, Wood BA, Tamariz L. Motivational interviewing improves medication adherence: a systematic review and meta-analysis. J Gen Intern Med 2016 Aug;31(8):929-940. [doi: 10.1007/s11606-016-3685-3] [Medline: 27160414]
37. Aliabadi N, Carballo-Dieguez A, Bakken S, Rojas M, Brown W, Carry M, et al. Using the information-motivation-behavioral skills model to guide the development of an HIV prevention smartphone application for high-risk MSM. AIDS Educ Prev 2015 Dec;27(6):522-537. [doi: 10.1521/aeap.2015.27.6.522] [Medline: 26595265]
38. Hawkins RP, Kreuter M, Resnicow K, Fishbein M, Dijkstra A. Understanding tailoring in communicating about health. Health Educ Res 2008 Jun;23(3):454-466 [FREE Full text] [doi: 10.1093/her/cyn004] [Medline: 18349033]
39. Anderson PL, Liu AY, Castillo-Mancilla JR, Gardner EM, Seifert SM, McHugh C, et al. Intracellular tenofovir-diphosphate and emtricitabine-triphosphate in dried blood spots following directly observed therapy. Antimicrob Agents Chemother 2018 Jan;62(1) [FREE Full text] [doi: 10.1128/AAC.01710-17] [Medline: 29038282]
40. Kapogiannis BG, Nelson RM, Siberry GK, Lee S, Hazra R. Advancing HIV biomedical prevention research for at-risk adolescents. J Acquir Immune Defic Syndr 2018 Dec 15;79(5):535-542 [FREE Full text] [doi: 10.1097/QAI.0000000000001853] [Medline: 30204722]
41. Brown W, Giguere R, Sheinfil A, Ibitoye M, Balan I, Ho T, et al. Challenges and solutions implementing an SMS text message-based survey CASI and adherence reminders in an international biomedical HIV PrEP study (MTN 017). J Biomed Inform 2018 Apr;80:78-86 [FREE Full text] [doi: 10.1016/j.jbi.2018.02.018] [Medline: 29501908]
42. Ratchakit-Nedsuwan R, Nedsuwan S, Sawadna V, Chaiyasirinroje B, Bupachat S, Ngamwithayapong-Yanai J, et al. Ensuring tuberculosis treatment adherence with a mobile-based CARE-call system in Thailand: a pilot study. Infect Dis (Lond) 2020 Feb;52(2):121-129. [doi: 10.1080/23744235.2019.1688862] [Medline: 31710250]
43. Phanuphak N, Anand T, Jantarapakde J, Nitpolprasert C, Himmad K, Sungsing T, et al. What would you choose: online or offline or mixed services? Feasibility of online HIV counselling and testing among Thai men who have sex with men and transgender women and factors associated with service uptake. J Int AIDS Soc 2018 Jul;21 Suppl 5:e25118 [FREE Full text] [doi: 10.1002/jia2.25118] [Medline: 30033644]
44. Cambon L. Health smart devices and applications…towards a new model of prevention? Eur J Public Health 2017;27(3):390-391. [doi: 10.1093/eurpub/ckx019]
45. Muessig KE, Pike EC, Legrand S, Hightow-Weidman LB. Mobile phone applications for the care and prevention of HIV and other sexually transmitted diseases: a review. J Med Internet Res 2013;15(1):e1 [FREE Full text] [doi: 10.2196/jmir.2301] [Medline: 23291245]
46. Aromatario O, Van Hoye A, Vuillemin A, Foucaut AM, Crozet C, Pommier J, et al. How do mobile health applications support behaviour changes? A scoping review of mobile health applications relating to physical activity and eating behaviours. Public Health 2019 Oct;175:8-18 [FREE Full text] [doi: 10.1016/j.puhe.2019.06.011] [Medline: 31374453]
47. Operario D, Kuo C, Sosa-Rubí SG, Gálarraga O. Conditional economic incentives for reducing HIV risk behaviors: integration of psychology and behavioral economics. Health Psychol 2013 Sep;32(9):932-940 [FREE Full text] [doi: 10.1037/a0032760] [Medline: 24001243]
48. Vlaev I, King D, Darzi A, Dolan P. Changing health behaviors using financial incentives: a review from behavioral economics. BMC Public Health 2019 Aug 07;19(1):1059 [FREE Full text] [doi: 10.1186/s12889-019-7407-8] [Medline: 31391010]
49. Ochonye B, Folayan MO, Fatusi AO, Emmanuel G, Adepoju O, Ajidagba B, et al. Satisfaction with use of public health and peer-led facilities for HIV prevention services by key populations in Nigeria. BMC Health Serv Res 2019 Nov 21;19(1):856 [FREE Full text] [doi: 10.1186/s12913-019-4691-z] [Medline: 31752853]
50. Iyengar K, Upadhyaya GK, Vaishya R, Jain V. COVID-19 and applications of smartphone technology in the current pandemic. Diabetes Metab Syndr 2020;14(5):733-737 [FREE Full text] [doi: 10.1016/j.dsx.2020.05.033] [Medline: 32497963]
51. Ming LC, Untong N, Aliudin NA, Osili N, Kifli N, Tan CS, et al. Mobile health apps on COVID-19 launched in the early days of the pandemic: content analysis and review. JMIR Mhealth Uhealth 2020 Sep 16;8(9):e19796 [FREE Full text] [doi: 10.2196/19796] [Medline: 32609622]
52. Lesko CR, Bengtson AM. HIV and COVID-19: intersecting epidemics with many unknowns. Am J Epidemiol 2021 Jan 04;190(1):10-16 [FREE Full text] [doi: 10.1093/aje/kwaa158] [Medline: 32696057]

53. Hoagland B, Torres TS, Bezerra DR, Geraldo K, Pimenta C, Veloso VG, et al. Telemedicine as a tool for PrEP delivery during the COVID-19 pandemic in a large HIV prevention service in Rio de Janeiro-Brazil. Braz J Infect Dis 2020 May 31 [FREE Full text] [doi: 10.1016/j.bjid.2020.05.004] [Medline: 32504552]
54. Sun Y, Li H, Luo G, Meng X, Guo W, Fitzpatrick T, et al. Antiretroviral treatment interruption among people living with HIV during COVID-19 outbreak in China: a nationwide cross-sectional study. J Int AIDS Soc 2020 Nov;23(11):e25637 [FREE Full text] [doi: 10.1002/jia2.25637] [Medline: 33247541]
55. Rao A, Rucinski K, Jarrett BA, Ackerman B, Wallach S, Marcus J, et al. Perceived interruptions to HIV prevention and treatment services associated with COVID-19 for gay, bisexual, and other men who have sex with men in 20 countries. J Acquir Immune Defic Syndr 2021 May 01;87(1):644-651. [doi: 10.1097/QAI.0000000000002620] [Medline: 33443963]

Abbreviations

CBO: community-based organization
CIPHER: Collaborative Initiative for Pediatric HIV Education and Research
DBS: dried blood spot
FGD: focus group discussion
IMB: information–motivation–behavioral
LRA: log-in and risk assessment
MSM: men who have sex with men
PrEP: pre-exposure prophylaxis
RSAT: Rainbow Sky Association of Thailand
SWING: Service Workers in Group
TFV-DP: tenofovir diphosphate
TGW: transgender women
TRCARC: Thai Red Cross AIDS Research Center
YFS: youth-friendly services
YFS+APP: youth-friendly services plus mobile phone app
YMSM: young men who have sex with men
YTGW: young transgender women

Edited by L Buis; submitted 07.11.20; peer-reviewed by A Queiroz, J Jones; comments to author 17.12.20; revised version received 11.02.21; accepted 09.12.21; published 21.04.22.

Please cite as:
Kawichai S, Songtaweesin WN, Wongharn P, Phanuphak N, Cressey TR, Moonwong J, Vasinonta A, Saisaengjan C, Chinbunchorn T, Puthanakit T
A Mobile Phone App to Support Adherence to Daily HIV Pre-exposure Prophylaxis Engagement Among Young Men Who Have Sex With Men and Transgender Women Aged 15 to 19 Years in Thailand: Pilot Randomized Controlled Trial
JMIR Mhealth Uhealth 2022;10(4):e25561
URL: https://mhealth.jmir.org/2022/4/e25561
doi:10.2196/25561
PMID:35451976

Original Paper

An Exercise and Educational and Self-management Program Delivered With a Smartphone App (CareHand) in Adults With Rheumatoid Arthritis of the Hands: Randomized Controlled Trial

Pablo Rodríguez Sánchez-Laulhé[1,2], PT, MSci; Luis Gabriel Luque-Romero[3,4], MD, PhD; Francisco José Barrero-García[3], MD; Ángela Biscarri-Carbonero[3], MSci, MD; Jesús Blanquero[1], PT, PhD; Alejandro Suero-Pineda[1], PT, MSci; Alberto Marcos Heredia-Rizo[1,2], PT, PhD

[1]Department of Physiotherapy, Faculty of Nursing, Physiotherapy and Podiatry, University of Seville, Seville, Spain

[2]Uncertainty, Mindfulness, Self, Spirituality (UMSS) Research Group, University of Seville, Seville, Spain

[3]Research Unit, Distrito Sanitario Aljarafe-Sevilla Norte, Andalusian Health Service, Seville, Spain

[4]Normal and Pathological Cytology and Histology Department, University of Seville, Seville, Spain

Corresponding Author:
Pablo Rodríguez Sánchez-Laulhé, PT, MSci
Department of Physiotherapy
Faculty of Nursing, Physiotherapy and Podiatry
University of Seville
Avicena St
Seville, 41009
Spain
Phone: 34 954486507
Email: pablorodriguez@healthinn.es

Abstract

Background: Rheumatoid arthritis (RA) is a prevalent autoimmune disease that usually involves problems of the hand or wrist. Current evidence recommends a multimodal therapy including exercise, self-management, and educational strategies. To date, the efficacy of this approach, as delivered using a smartphone app, has been scarcely investigated.

Objective: This study aims to assess the short- and medium-term efficacy of a digital app (CareHand) that includes a tailored home exercise program, together with educational and self-management recommendations, compared with usual care, for people with RA of the hands.

Methods: A single-blinded randomized controlled trial was conducted between March 2020 and February 2021, including 36 participants with RA of the hands (women: 22/36, 61%) from 2 community health care centers. Participants were allocated to use the CareHand app, consisting of tailored exercise programs, and self-management and monitoring tools or to a control group that received a written home exercise routine and recommendations, as per the usual protocol provided at primary care settings. Both interventions lasted for 3 months (4 times a week). The primary outcome was hand function, assessed using the Michigan Hand Outcome Questionnaire (MHQ). Secondary measures included pain and stiffness intensity (visual analog scale), grip strength (dynamometer), pinch strength (pinch gauge), and upper limb function (shortened version of the Disabilities of the Arm, Shoulder, and Hand questionnaire). All measures were collected at baseline and at a 3-month follow-up. Furthermore, the MHQ and self-reported stiffness were assessed 6 months after baseline, whereas pain intensity and scores on the shortened version of the Disabilities of the Arm, Shoulder, and Hand questionnaire were collected at the 1-, 3-, and 6-month follow-ups.

Results: In total, 30 individuals, corresponding to 58 hands (CareHand group: 26/58, 45%; control group: 32/58, 55%), were included in the analysis; 53% (19/36) of the participants received disease-modifying antirheumatic drug treatment. The ANOVA demonstrated a significant time×group effect for the total score of the MHQ ($F_{1.62,85.67}$=9.163; P<.001; η^2=0.15) and for several of its subscales: overall hand function, work performance, pain, and satisfaction (all P<.05), with mean differences between groups for the total score of 16.86 points (95% CI 8.70-25.03) at 3 months and 17.21 points (95% CI 4.78-29.63) at 6 months. No time×group interaction was observed for the secondary measures (all P>.05).

Conclusions: Adults with RA of the hands who used the CareHand app reported better results in the short and medium term for overall hand function, work performance, pain, and satisfaction, compared with usual care. The findings of this study suggest

that the CareHand app is a promising tool for delivering exercise therapy and self-management recommendations to this population. Results must be interpreted with caution because of the lack of efficacy of the secondary outcomes.

Trial Registration: ClinicalTrials.gov NCT04263974; https://clinicaltrials.gov/ct2/show/NCT04263974

International Registered Report Identifier (IRRID): RR2-10.1186/s13063-020-04713-4

(JMIR Mhealth Uhealth 2022;10(4):e35462) doi:10.2196/35462

KEYWORDS

rheumatoid arthritis; telerehabilitation; self-management; mHealth; primary health care; physical therapy; exercise therapy; mobile applications; telehealth; health education; mobile phone

Introduction

Background

Rheumatoid arthritis (RA) is one of the most frequent systemic autoimmune diseases globally (approximately 1% of the population worldwide), with a higher prevalence in women [1]. In Spain, up to 430,000 adults aged >20 years have been estimated to have this disease [2]. RA results in tissue damage and chronic inflammation [3], especially in the small synovial joints of the hands and wrists [4]. Clinical presentation often involves musculoskeletal deficits; for example, hand deformities, pain, and reduced grip strength (GS) and pinch strength [5], which lead to functional and social limitations [6], along with a decline in work ability, productivity losses [7], and worse quality of life [8]. Together, these factors cause a substantial socioeconomic burden [9].

Pharmacological management of RA with disease-modifying antirheumatic drugs (DMARDs) and nonsteroidal anti-inflammatory drugs can help decrease RA-related symptoms and progression [8], although it may also induce serious adverse events [8] and largely increase health care costs [10]. Therefore, beneficial, safe, and cost-effective interventions must be implemented and prioritized in daily settings. As such, exercise therapy, supervised or at home, has been proposed as a suitable first-line approach for people with RA [10-12]. The current literature suggests, with inconclusive evidence [11], that exercise training programs for RA of the hands may improve the range of movement [13] and hand and upper limb functions [14,15]. In addition, they can reduce muscle weakness [16], pain intensity [17], and disease flare-ups [18] and eventually enhance the effects of DMARDs [15].

Long-term adherence to treatment is challenging and typically low in patients with RA [19]. This increases the risk of higher disability [20] and compromises the efficacy of the therapy [21]. Implementing strategies to solve this issue appears to be essential [11]. Several recommendations have been made, including an exercise diary to foster self-management and monitor treatment progression; for example, dose and intensity of exercises, establishing realistic therapy goals, a verbal or written commitment from the patient [12], providing information and education about the disease [12,19], and maintaining regular email or telephone contact [22], among others.

eHealth, as the use of communication technologies to support health-related fields, is a feasible solution for intervention delivery and has become imperative for health care systems, even more in the current COVID-19 pandemic context [23,24]. Mobile apps are the most effective eHealth modalities for reducing pain interference in chronic pain conditions [25]. Similarly, telehealth exercise programs have been shown to be an alternative to treat pain, physical function, and quality of life in people with physical disabilities [26,27]. Recent literature concludes that feedback-guided exercises delivered with a tablet increase function after carpal tunnel release [28] and hasten return to work in individuals with wrist, hand, or finger injuries [29]. In adults with RA, digital interventions using smartphone apps are promising to support clinical care and empower self-management [30], although they may not be effective unless designed including evidence-based strategies to promote adherence [31]. Preliminary findings demonstrate that digital apps that encourage self-management and allow for self-monitoring of the condition can improve health outcomes in patients with RA of the hands [32,33].

The CareHand app (Healthinn) was designed and developed with input from users and under the supervision of experts to meet the latest scientific evidence and the needs of patients and professionals, which is uncommon in existing mobile health apps [34]. This app emerges as a solution for telerehabilitation of patients with rheumatic hands and includes strategies to foster active self-management routines and long-term adherence, but it needs to be evaluated in the clinical setting.

Objectives

This study aims to investigate the short- and medium-term effectiveness of a home therapeutic exercise program combined with general and self-management recommendations, as implemented with a mobile app (CareHand), compared with a usual care approach (exercise program and recommendations on a paper sheet) in people with RA of the hands. We hypothesized that hand function would improve more for participants who used the CareHand app.

Methods

Study Design

An experimental, longitudinal, parallel, controlled, and single-blinded randomized trial was conducted following the published protocol [35], prospectively registered at ClinicalTrials.gov (NCT04263974) and including an extended long-term follow-up of 6 months.

Ethics Approval

The research protocol complied with the ethical guidelines of the Declaration of Helsinki and was approved by the Research

Ethics Committee of the Virgen del Rocio and Virgen Macarena University Hospitals, Seville, Spain (code number PI_RH_2018).

Participants

Participants aged ≥18 years and with a medical diagnosis of RA of the hands, wrists, or fingers, based on the American College of Rheumatology guidelines [36], were selected through data available from the digital medical records at the Health Districts Northern Seville and Aljarafe, Andalusian Health Service, Seville, Spain. Eligible individuals were contacted via telephone and asked to participate.

To be included in the study, participants had to have a disease history lasting for at least two years [36], report current pain and disability in the hands or wrists [37], and possess a smartphone with internet access. Exclusion criteria were previous hand fracture or surgery [37], waiting for upper limb surgery [15], steroid injection in the month before recruitment [19], pregnancy [38], and diagnosis of cognitive problems that may preclude the completion of the study protocol [39].

Intervention Strategies

Participants in both groups were asked to perform their exercise intervention protocol at home 4 times a week for 3 months, with each training session programmed to last approximately 15 to 20 minutes. Several telephone follow-up calls were made during the trial to monitor adherence to the intervention and to solve eventual problems.

The control group underwent the conventional primary care approach of the public health system where the study was conducted. This consisted of providing a written exercise program and recommendations on a paper sheet, together with pictures and written explanations of upper limb strengthening and stretching exercises focusing on the hands, wrist, and finger joints.

Individuals in the CareHand group were asked to use the CareHand app. This digital solution has been developed under the guidance of health care professionals (physiotherapists and physicians) for its use on Android or iOS smartphones and comprises treatment and monitoring systems for the rehabilitation of people with rheumatic hands. The company responsible for the app has successfully implemented a tablet application (ReHand) for the treatment of traumatic injuries of the wrist, hand, and fingers [28,29], but no studies have investigated the efficacy of CareHand.

The training program of the app, based on clinical guidelines for exercise therapy in people with RA of the hands [38,40], was delivered with explanatory videos, including a warm-up routine, and mobility, stretching, and strengthening exercises (Figure 1). Users of the app had to report their pain intensity twice a day (before and after the exercise program) and respond to self-reported outcome questionnaires once a week. The load and intensity of the exercises were automatically adapted to each participant's pain intensity [41-43] (Figure 2). The CareHand app includes an exercise diary with a graphical representation of the progress in the planned treatment protocol and the evolution of pain intensity. This diary helps to promote positive self-management routines, monitor long-term adherence, and collect patient feedback [12,38]. The app also provides advice on diet and rules for joint protection. Educational and self-management strategies to handle RA-related symptoms and improve function during activities of daily living (ADL), along with recommendations for regular physical activity, were also provided (Figure 2) [19,38]. The app recorded and sent adherence charts to a cloud database. The data were monitored by the team member in charge of the interventions to ensure proper compliance with treatment.

Figure 1. Exercise program of the CareHand app, including explanatory videos of mobility, strengthening, and stretching exercises.

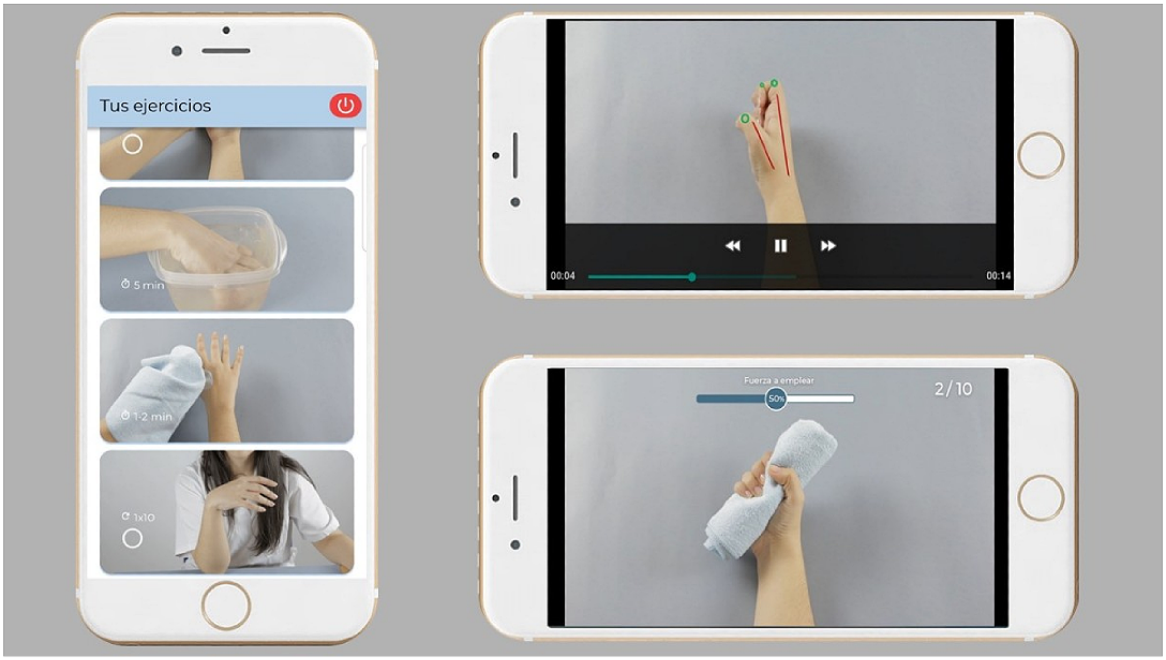
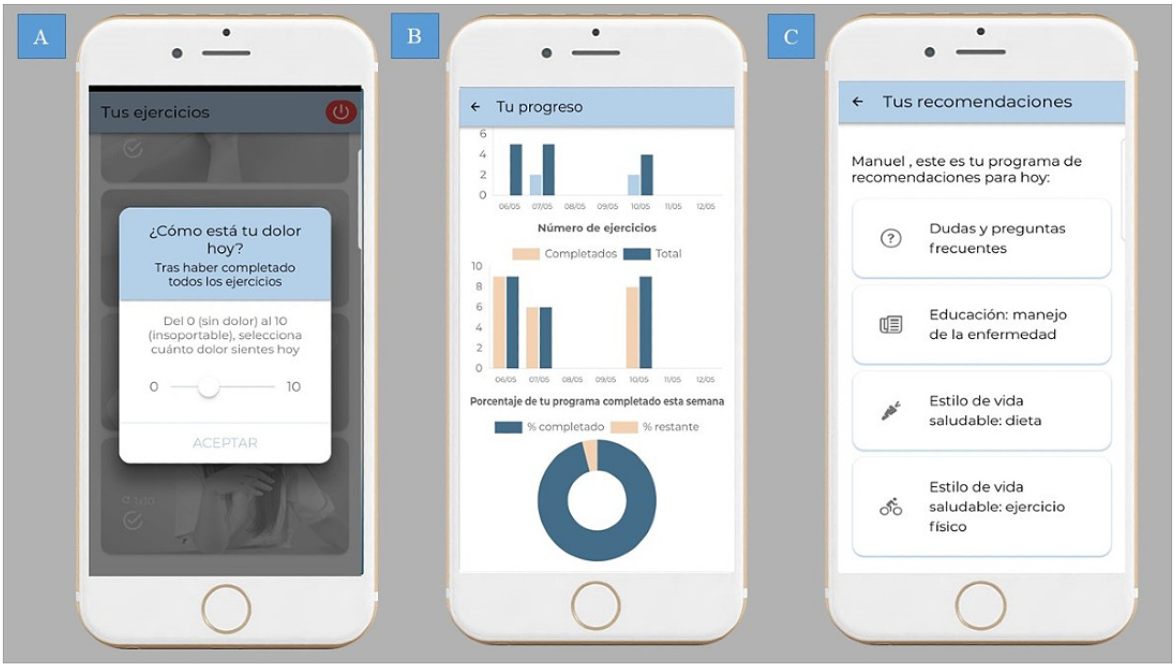


Figure 2. Features of the CareHand app. (A) Self-monitoring for pain intensity, (B) graphical representation of patient progress and adherence to exercises, and (C) educational advices section with information about joint protection and general recommendations.



Enrollment

After contacted by telephone, those interested were scheduled for a face-to-face session at a community health center located at Camas or Sanlúcar la Mayor, Seville, Spain. During this session, a researcher (PRSL) assessed whether the individuals fulfilled the eligibility criteria. After agreeing to enroll, participants received further information about the trial and were asked to provide written informed consent. Then, clinical and demographic data were collected. After that, the participant was walked to a different room, and another assessor (LGLR, FJBG, or ABC), who was a general practitioner previously trained to evaluate the study measures, collected the outcomes at baseline: hand and upper limb function, self-reported pain intensity and stiffness, and GS and pinch strength. Following the baseline assessment, patients were randomly assigned to a study group using a random sequence in permuted blocks to allow a 1:1 distribution ratio. Sealed opaque envelopes were used to conceal intervention allocation. Participants were scheduled for a second appointment a week later, where the lead researcher (PRSL) explained the training protocol. This informative appointment was the starting session of the intervention.

Outcome Measurements

Overview

Different measures were used to evaluate the efficacy of the interventions, including the Michigan Hand Outcome Questionnaire (MHQ) [44] for hand function; a visual analog scale (VAS) for self-reported pain and morning stiffness intensity [45,46]; a hydraulic hand dynamometer for GS; a pinch gauge for pinch strength [47]; and the shortened version of the Disabilities of the Arm, Shoulder, and Hand questionnaire (QuickDASH) [48] to measure upper limb function.

The lead investigator (PRSL) collected demographic and personal data at baseline, including age, gender, DMARDs consumption, and dominant hand.

Primary Outcome: Hand Function

The MHQ is an appropriate tool for individuals with chronic conditions of the hand. The questionnaire is divided into six subscales: function, ADL, work performance, pain, esthetic, and satisfaction [49]. Final scores ranged from 0 to 100, with higher values denoting better hand performance, except for pain [50]. The MHQ has shown high validity, reliability, and sensitivity in people with rheumatoid hands [44,50]. The Spanish version of the MHQ has good validity, reliability, and sensitivity to change [51]. This outcome was collected at baseline and at the 3- and 6-month follow-ups.

Secondary Outcomes

Participants reported their average pain and morning stiffness intensity in the previous week using a 11-point VAS. This measure shows good psychometric values when used to assess self-reported stiffness in patients with RA [45] and pain intensity in people with hand disorders [46]. Pain intensity was collected four times: at baseline and at the 1-, 3-, and 6-month follow-ups, whereas stiffness was collected at baseline and at the 3- and 6-month follow-ups.

A hand dynamometer (Saehan SH5001, Saehan Corp) was used, following the American Society of Hand Therapy statements, to evaluate GS [47]. Reporting GS is easy, quick, and reliable [52] and appears to be strongly related to the level of disability of rheumatic hands [53]. This procedure has demonstrated a great test-retest reliability; thus, measures were taken only once to avoid patient discomfort [54]. Maximum pain-free pinch force was assessed with a pinch gauge (13.5-kg mechanical pinch gauge, Baseline) [47], using a single measurement [55].

Pinch strength is inversely related to hand and upper limb function in patients with RA [56]. We collected GS and pinch strength data at baseline and at 3 months after the intervention.

Regarding upper limb function, the QuickDASH is a valid, reliable, easy to use, and widely used tool in patients with RA [5,48,57]. The questionnaire was completed at baseline and at the 1-, 3-, and 6-month follow-ups.

Statistical Analysis

The sample size was calculated to achieve clinically significant differences (<13 points) [58], with a medium effect size ($0.06 \leq \eta^2 \leq 0.14$), for the overall hand function of the MHQ in the comparison between groups after intervention. RA is a symmetrical disease that involves joints bilaterally in over 60% of patients, although symptoms may [4] differ between sides [4]. Therefore, although the protocol estimation was made in terms of participants [35], we decided to deviate from the initial protocol and report the sample size in terms of the number of hands treated to adhere to common clinical practice. For an 80% desired power, an α value of .05, a correlation among repeated measures of 0.5, and a within-group variance of 10, a total of 56 hands were needed to complete the study (G*Power software, version 3.1.9.7; Kiel University).

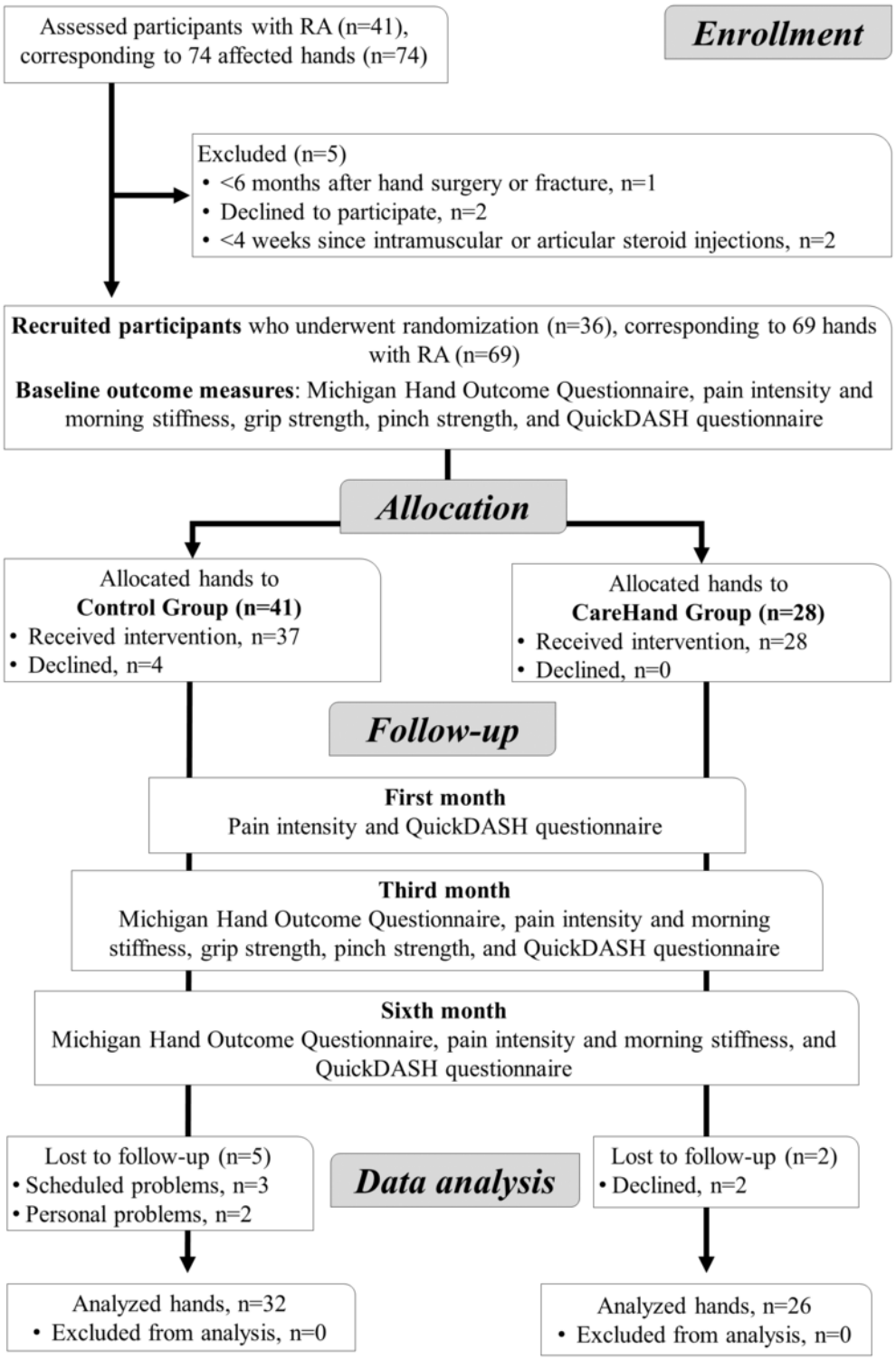
Intention-to-treat principles were considered for all statistical analyses, which were conducted using SPSS Statistics (version 26; IBM Corp) software. Data are reported as mean (SD or 95% CI) or as percentages. The Shapiro-Wilk test was used to evaluate the normal distribution of the study measures. Mean outcome differences after intervention were compared using repeated-measures ANOVA, with group (control vs CareHand) as the between-subjects factor and time (baseline and the first-, third-, and sixth-month follow-ups) as the within-subjects factor. When the assumption of sphericity was violated, the Greenhouse-Heisser correction was applied. Partial eta squared values estimate the effect size. Statistical significance was set at $P<.05$.

Results

Flow of Participants

Between March 2020 and February 2021, a total of 41 adults with unilateral or bilateral RA of the hand were recruited. In those with a bilateral condition, both hands were selected if the inclusion and exclusion criteria were fulfilled. Of all 74 rheumatic hands, 5 (7%) were excluded at baseline, with 69 (93%) hands eligible for the study. Moreover, of the 36 recruited participants, 6 (17%) dropped out after baseline. Finally, 30 (83%) adults, corresponding to 58 hands (n=58), were included for statistical analysis. The flowchart of the participants is shown in Figure 3.

Figure 3. CONSORT (Consolidated Standards of Reporting Trials) flowchart of participants. QuickDASH: shortened version of the Disabilities of the Arm, Shoulder, and Hand questionnaire; RA: rheumatoid arthritis.



Participant Characteristics

The baseline clinical and demographic characteristics of the participants are presented in Table 1. The patients were aged between 43 and 78 years, including 61% (22/36) women. A total of 92% (33/36) participants were right-handed, and 53% (19/36) participants were under pharmacological treatment with DMARDs. Moreover, 52% (36/69) of the hands that underwent intervention were right hands. Before the intervention, there were differences between the groups in the pain (P=.001) and satisfaction domains (P=.03) of the MHQ.

Table 1. Clinical and demographic baseline characteristics of participants.

	CareHand group (n=14, 28 hands)	Control group (n=22, 41 hands)	*P* value
Age (years), mean (SD)	57.64 (7.25)	61.86 (10.76)	.21
Gender (female), n (%)	9 (64)	13 (59)	.76
DMARD[a] treatment (yes), n (%)	10 (71)	9 (40)	.08
Dominant hand, n (%)			.71
Right	13 (93)	20 (91)	
Left	0 (0)	2 (9)	
Both	1 (7)	0 (0)	
Affected hand (right), n (%)	14 (50)	22 (54)	.77
MHQ,[b] mean (SD)			
Overall hand function	50.89 (12.84)	57.81 (20.03)	.11
Activities of daily living	61.45 (22.94)	63.50 (31.66)	.77
Work performance	45 (29.25)	53.25 (35.47)	.32
Pain	62.50 (22.42)	39.02 (28.82)	.001
Esthetics	69.42 (28.28)	79.42 (29.32)	.16
Satisfaction	39.14 (21.41)	52.13 (25.38)	.03
Total score	50.57 (18.46)	61.09 (23.45)	.051
Pain (VAS[c]; 0-10), mean (SD)	4.84 (2.76)	4.73 (2.79)	.87
Stiffness (VAS; 0-10), mean (SD)	4.54 (2.86)	5.31 (3.41)	.33
Grip strength (kg), mean (SD)	14.37 (7.91)	16.74 (9.65)	.29
Pinch strength (kg), mean (SD)	4.15 (1.80)	4.52 (1.91)	.43
QuickDASH[d] (0-100), mean (SD)	52.27 (15.36)	43.05 (28.86)	.28

[a]DMARD: disease-modifying antirheumatic drug.

[b]MHQ: Michigan Hand Outcome Questionnaire.

[c]VAS: visual analog scale.

[d]QuickDASH: shortened version of the Disabilities of the Arm, Shoulder, and Hand questionnaire.

Hand Function

For the primary outcome, the ANOVA demonstrated a significant time×group effect for the total score of the MHQ ($F_{1.62,85.67}$=9.163; P<.001; η^2=0.15), with mean differences between groups of 16.86 points (95% CI 8.70-25.03) at the third-month follow-up and 17.21 points (95% CI 4.78-29.63) at the sixth-month follow-up. Furthermore, a statistically significant time×group effect was observed for several subscales of the MHQ: overall hand function ($F_{2,106}$=3.298; P=.04; η^2=0.06), work performance ($F_{2,98}$=6.892; P=.002; η^2=0.12), pain ($F_{2,106}$=13.918; P<.001; η^2=0.21), and satisfaction ($F_{1.69,89.77}$=5.949; P=.006; η^2=0.10; Table 2). The graphical representation of the mean differences in the MHQ in the 2 groups and in the different assessment points is included in Multimedia Appendix 1.

Table 2. Within- and between-group differences for the Michigan Hand Outcome Questionnaire (MHQ).

MHQ subscales	Within-groups differences from baseline, mean difference (95% CI)		Differences between CareHand and control groups		Time-group effect, *P* value
	CareHand group	Control group	Values, mean difference (95% CI)	*P* value	
Overall hand function					.04
Third month	11.67 (5.18 to 18.16)	−1.69 (−9.69 to 6.31)	13.36 (2.85 to 23.87)	.01	
Sixth month	7.50 (−1.82 to 16.82)	0.31 (−7.72 to 8.35)	7.19 (−4.79 to 19.16)	.23	
Activities of daily living					.26
Third month	11.55 (3.85 to 19.25)	5.59 (−2.70 to 13.87)	5.66 (−5.38 to 17.31)	.30	
Sixth month	9.83 (2.12 to 17.55)	3.19 (−5.14 to 11.52)	6.64 (−4.68 to 17.97)	.25	
Work performance					.006
Third month	10.00 (−2.94 to 22.94)	−0.97 (−10.94 to 9.00)	10.97 (−4.70 to 26.64)	.17	
Sixth month	18.33 (2.40 to 34.26)	−5.50 (−16.12 to 5.12)	23.83 (5.77 to 41.90)	.01	
Pain[a]					<.001
Third month	−22.50 (−33.5 to −11.49)	12.58 (1.59 to 23.57)	−35.08 (−50.54 to −19.62)	<.001	
Sixth month	−17.31 (−27.32 to −7.30)	8.75 (−0.82 to 18.32)	−26.06 (−39.69 to −12.42)	<.001	
Esthetic					.10
Third month	−0.78 (−12.77 to 11.21)	−11.28 (−22.80 to 0.23)	10.50 (5.93 to 26.93)	.21	
Sixth month	2.40 (−16.88 to 21.68)	−18.16 (−32.38 to −3.93)	20.56 (−2.37 to 43.48)	.08	
Satisfaction					.006
Third month	20.14 (10.33 to 29.96)	−3.23 (−10.93 to 4.47)	23.37 (11.37 to 35.36)	<.001	
Sixth month	14.59 (0.30 to 28.88)	−4.04 (−14.02 to 5.94)	18.62 (2.04 to 35.20)	.02	
Total					<.001
Third month	12.51 (5.48 to 19.55)	−4.35 (−9.32 to 0.62)	16.86 (8.70 to 25.03)	.001	
Sixth month	11.56 (1.88 to 21.24)	−5.65 (−13.97 to 2.68)	17.21 (4.78 to 29.63)	.007	

[a]In the MHQ pain subscale, higher scores represent worse pain status.

Secondary Outcomes

Scores for the secondary measures are presented in Table 3. The ANOVA reported no time×group interaction for any of the following outcomes: pain intensity ($F_{3,153}$=1.352; P=.26; η^2=0.03), morning stiffness ($F_{2,106}$=1.299; P=.28; η^2=0.02), GS ($F_{1,35}$=0.001; P=.99; η^2=0.001) and pinch strength ($F_{1,35}$=0.112; P=.74; η^2=0.003), and the QuickDASH ($F_{3,75}$=0.924; P=.43; η^2=0.04).

Table 3. Within- and between-group differences for self-reported pain and stiffness, grip strength and pinch strength, and upper limb function.

	Within-groups differences from baseline, mean difference (95% CI)		Differences between CareHand and control groups		Time-group effect, *P* value
	CareHand group	Control group	Values, mean difference (95% CI)	*P* value	
Pain intensity (VAS[a])					.26
First month	0.94 (−0.36 to 2.25)	−0.20 (−1.45 to 1.06)	1.14 (−0.64 to 2.91)	.20	
Third month	−0.44 (−1.42 to 0.54)	−0.32 (−1.39 to −0.75)	−0.12 (−1.58 to 1.34)	.87	
Sixth month	0.90 (−0.43 to 2.24)	0.50 (−0.67 to 1.68)	0.40 (−1.33 to 2.13)	.65	
Stiffness intensity (VAS)					.28
Third month	−0.38 (−1.89 to 1.14)	−0.95 (−1.81 to −0.09)	0.57 (−1.03 to 2.19)	.65	
Sixth month	0.19 (−1.31 to 1.69)	−1.02 (−2.30 to 0.27)	1.21 (−0.71 to 3.13)	.21	
Grip strength (kg)					
Third month	1.69 (−0.76 to 4.15)	1.69 (−0.35 to 3.73)	0.00 (−3.05 to 3.06)	.99	.99
Pinch strength (kg)					
Third month	−0.42 (−1.48 to 0.65)	−0.22 (−0.92 to 0.48)	−0.20 (−1.37 to 0.99)	.74	.74
QuickDASH[b]					.43
First month	−8.92 (−22.00 to 4.17)	−5.78 (−16.74 to 5.18)	−3.14 (−19.27 to 13.00)	.69	
Third month	−10.80 (−22.47 to 0.87)	−7.05 (−15.75 to 1.63)	−3.74 (−17.25 to 9.77)	.57	
Sixth month	−17.22 (−29.82 to −4.62)	−6.91 (−17.37 to 3.54)	−10.31 (−25.82 to 5.21)	.18	

[a]VAS: visual analog scale.

[b]QuickDASH: shortened version of the Disabilities of the Arm, Shoulder, and Hand questionnaire.

Adverse Effects

No adverse events related to interventions were reported throughout the trial. A total of 17% (6/36) of participants experienced disease-associated pain flare-ups during follow-up.

Discussion

Principal Findings

As hypothesized, the findings of this study suggest that the CareHand app was better than conventional care in improving hand functional ability (overall hand function, work performance, pain, and satisfaction) in the short and medium term in adults with RA of the hands. However, no differences between groups were demonstrated for self-reported pain and morning stiffness, GS and pinch strength, and upper limb function; thus, the results must be interpreted cautiously.

Hand Function

For the MHQ, the ANOVA showed a medium to large size effect in favor of the digital app. Mean differences between groups surpassed the clinically relevant thresholds for overall hand function (13 points) at the 3-month follow-up and for pain (11 points) at the 3- and 6-month follow-ups [58]. Pain and function are the best domains of the MHQ for identifying satisfied patients after treatment [58]. Recent evidence highlights the importance of considering hand function as the primary outcome in clinical trials investigating the efficacy of exercise therapy for RA of the hands [11], which should become a priority in daily practice [14].

The current literature on the impact of exercise on rheumatic hands is conflicting [11], although positive effects have mostly been reported [59]. Our findings are in line with those of a core trial on the topic, the Strengthening and Stretching for Rheumatoid Arthritis of the Hand (SARAH) project [15]. Lamb et al [15] concluded a positive effect on hand function (mean change: approximately 4.5 points), ADL, and total MHQ scores at 4 and 12 months after a multimodal approach with individualized exercises, strategies to enhance adherence, general recommendations, and joint protection education, as adjunct to drug treatment [15]. They also reported long-term improvements in hand dexterity. However, all observed changes only suggested a minimal clinical impact [15]. This study conducted an extended follow-up beyond 2 years, when the efficacy of the program diminished considerably compared with usual care [60]. Good results for hand function have also been reported when combining active hand exercises with wax baths [61]. In contrast, adding hand strengthening and mobility exercises to joint protection information was not superior to information alone in enhancing hand and finger function at 6 months [14]. Similarly, an 8-week exercise program, together with compensatory strategies; for example, joint protection, and use of assistive devices, added no benefits for task performance or ADL ability in women with RA of the hands [37]. These contradictory findings, along with the low quality of most trials on the topic, warrant new research [11].

eHealth has become an alternative and cost-saving approach to make evidence-based treatments available for patients and clinicians and to foster proactive self-management [30]. In this

context, a web-based self-guided exercise program has been tested in adults with RA of the hands [62], with promising preliminary findings for hand function at 12 and 16 weeks [63]. Different mobile apps are also used by people with RA. However, most of them provide either symptom tracking or information alone and lack a comprehensive experience for patients [64]. Very recently, a mobile app including a structured hand exercise program was assessed for usability in adults with rheumatic hands, with good levels of satisfaction [65]. To date, the CareHand app is the first to be investigated in a clinical setting. This app includes educational information and adherence strategies, together with symptom tracking (exercise diary) through attractive audiovisual material [35]. The CareHand app also individualizes the exercise dose based on the self-reported pain intensity. This multimodal approach has proven to be effective in increasing hand function and self-efficacy in this population [15,19]. Higher self-efficacy is an empowerment protective factor in people with chronic symptoms and has been correlated with better hand function and quality of life in adults with RA [66]. All these app features may help to increase adherence to intervention [19] and treatment efficacy [11], even in the long run [21], which would explain our positive findings for the primary measure.

Secondary Measures

According to the present evidence, systemic exercise treatment may be suitable for reducing RA-related pain [59]. However, it is still uncertain whether upper limb exercise therapy decreases hand pain or stiffness in adults with rheumatic hands [11]. In our study, the findings showed no effect of any of the interventions on self-reported pain or morning stiffness. Our results agree with those of previous trials where home exercise programs, alone or together with education and self-management routines, did not change pain intensity in the medium term [15,37,67,68] or long term [15,60] in this population. Similar findings were obtained when using digital technology to deliver an exercise regime [63]. In contrast, other trials have found that combining exercise with other forms of physical therapy helps to reduce hand pain, both in the short [16] and medium term [19,39,61,69]. RA is a condition that usually involves periods of flare-ups, which could affect self-reported pain scores and explain the conflicting evidence on the topic [70]. In addition, more than half of the participants (19/36, 53%) used DMARDs, which could also have influenced this outcome. Unsurprisingly, pain intensity did not improve despite changes in hand function. However, pain and function are not necessarily associated factors. This has been demonstrated in individuals with RA [37] and, most importantly, in people with chronic conditions [71] probably because of the multifactorial etiology of persistent pain [71]. Therefore, although therapists and patients with RA usually consider pain intensity as an important clinical measure [65], a recent systematic review proposed the use of function instead of pain as the primary outcome in this population [11]. Morning stiffness is one of the first symptoms caused by RA and is a predictor of a poor prognosis [45]. With regard to stiffness, evidence from the scientific literature is also contradictory, as observed for pain. Some studies have found good results for this outcome after exercise training [61], although most evidence points out the lack of efficacy of exercise programs for reducing morning hand stiffness [11,19,72]. The lack of treatment responsiveness has been partly explained by the great variability of this symptom within and between patients with RA [73].

Clinical and research guidelines recommend assessing GS and pinch strength in people with RA or osteoarthritis of the hands [74], as lower hand strength could be related to reduced functional ability [56] and greater structural joint damage [53]. The present literature is conflicting and unclear regarding this issue. Overall, very low–quality evidence indicates that hand exercise training, compared with no treatment, may improve GS and pinch strength in the short term [11]. However, when compared with usual care, it seems to have little or no benefit on hand strength in people with RA of the hands [11]. This could be a plausible explanation for our results. In line with this, joint protection programs [75] or general aerobic exercises [76] are not effective in increasing GS in patients with hand arthritis. Given the course of the disease and the heterogeneity of training protocols among studies, it is difficult to reach a definite conclusion. In addition, there are many different person-related factors; for example, age, gender, dominant hand, work occupation, leisure activities, and psychological aspects, that may influence this outcome in people with arthritic conditions, whether RA [53] or osteoarthritis [74]. Future studies should control for these confounding variables [53].

Finally, upper limb function is an important measure in RA, as it is associated with disease activity [19,77], self-efficacy [19], sensorimotor deficits [78], and quality of life [77]. However, there is little evidence regarding the effect of exercise training on this outcome. In the within-group analysis, we found improvements in both groups that surpassed the smallest detectable difference for the QuickDASH (6.9 points) [79] in the short and medium term. This may help to explain the lack of time×group effect, although differences between groups were clinically relevant at 6 months (−10.31 points, 95% CI −25.82 to 5.21 points). However, it has been questioned whether the QuickDASH is specific enough for people with RA of the hands [80]. Among the scarce studies in this area, Manning et al [19] delivered a similar intervention, using an Education, Self-Management, and Upper Extremity Exercise Training (EXTRA) program, with positive results for upper limb function at 12 weeks, compared with usual care, which eventually disappeared at 36 weeks, in line with former trials [68]. The scant evidence suggests a potential benefit of exercise therapy in decreasing upper limb disability, but further research is needed to support this statement.

Limitations

This study has several limitations. The sample size was rather small, although relevant for the study aim and clinical purposes. In addition, there was a deviation from the initial protocol in terms of reporting the sample size estimation. This was intended to adhere pragmatically to the common standard practice of the clinical setting where the study was conducted. Despite the multicenter design, the participants were selected from a rural setting, which could limit the external validity of the findings. Changes in the study measures were evaluated in a medium-term follow-up (6 months); however, a long-term assessment is

needed to better understand the efficacy of digital tools and the impact of the strategies used to engage patients. Important factors such as self-efficacy, psychological aspects, and treatment adherence were not measured. At baseline, both groups differed in the pain and satisfaction subscales of the MHQ, which may be a source of bias. In addition, owing to the restrictions imposed by the COVID-19 pandemic, some of the outcomes were collected via telephone. Finally, the normal course of RA includes periods of remission and exacerbation of symptoms, which can affect self-reported data.

Clinical Implications

The CareHand app is a usable digital tool that opens a new field for the management of chronic rheumatic conditions of the hand. This app, developed with feedback from patients and health professionals, allows clinicians to treat people with RA and monitor their symptoms, evolution, and engagement with intervention. The app features foster proactive self-management. This may enhance self-efficacy and empower patients, which is key to managing chronic musculoskeletal disorders [81]. The feedback features also allow for a quick response if a disease flare-up appears [82]. The wide use of mobile devices and their portability represent a great potential impact of this app on health care delivery processes [65]. When implemented in a clinical setting, the CareHand app could reduce unnecessary visits to medical centers, as observed with other telehealth strategies in RA [83], with subsequent economic implications.

Conclusions

A multimodal approach including a home exercise regime and self-management recommendations, as delivered with the CareHand smartphone app, was more effective than providing written instructions for the exercise program to improve hand function in the short and medium term in adults with RA of the hands. Despite these promising findings, no effects were found for self-reported pain intensity and stiffness, hand strength, and upper limb function.

Acknowledgments

This study is part of a larger project titled *CareHand* that has received financial support from the Fundación Pública Andaluza Progreso y Salud, Consejería de Salud y Familias, Junta de Andalucía, Spain (reference number AP-0149-2017). The authors would like to thank Vicente Rodríguez-Pappalardo, Antonio García-López, and the public health centers at Camas and Sanlúcar la Mayor, Seville, Spain, for their support during recruitment and data collection and the engineering (María de las Nieves Sánchez-Laulhé and José Manuel López) and quality (Lorena Sarmiento) teams responsible for the development and quality and safety evaluation of the CareHand app.

Authors' Contributions

LGLR is the principal investigator and coordinated the study. PRSL, LGLR, ASP, and JB designed the study. PRSL and LGLR supervised the appointment dates of the participants and coordinated the research team. PRSL was responsible for the allocation and conducted patient training sessions. PRSL, LGLR, FJBG, and ABC carried out data collection and telephone monitoring of patients. PRSL, AMHR, ASP, and JB were the lead contributors for this manuscript. This final version has been reviewed and approved by all authors.

Conflicts of Interest

JB, ASP, and PRSL have been members of the ReHand project since 2016 and are part of the spin-off (Healthinn) that has designed and developed the CareHand app.

Multimedia Appendix 1
Mean changes from baseline (A) to 3-month (B) and 6-month (C) follow-ups for the Michigan Hand Outcome Questionnaire (total and subscale scores).
[PDF File (Adobe PDF File), 76 KB - mhealth_v10i4e35462_app1.pdf]

Multimedia Appendix 2
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1288 KB - mhealth_v10i4e35462_app2.pdf]

References

1. Alamanos Y, Voulgari PV, Drosos AA. Incidence and prevalence of rheumatoid arthritis, based on the 1987 American College of Rheumatology criteria: a systematic review. Semin Arthritis Rheum 2006;36(3):182-188. [doi: 10.1016/j.semarthrit.2006.08.006] [Medline: 17045630]
2. Silva-Fernández L, Macía-Villa C, Seoane-Mato D, Cortés-Verdú R, Romero-Pérez A, Quevedo-Vila V, et al. The prevalence of rheumatoid arthritis in Spain. Sci Rep 2020;10(1):21551 [FREE Full text] [doi: 10.1038/s41598-020-76511-6] [Medline: 33299019]

3. Zamanpoor M. The genetic pathogenesis, diagnosis and therapeutic insight of rheumatoid arthritis. Clin Genet 2019;95(5):547-557. [doi: 10.1111/cge.13498] [Medline: 30578544]
4. Aletaha D, Smolen JS. Diagnosis and management of rheumatoid arthritis: a review. JAMA 2018;320(13):1360-1372. [doi: 10.1001/jama.2018.13103] [Medline: 30285183]
5. Palamar D, Er G, Terlemez R, Ustun I, Can G, Saridogan M. Disease activity, handgrip strengths, and hand dexterity in patients with rheumatoid arthritis. Clin Rheumatol 2017;36(10):2201-2208. [doi: 10.1007/s10067-017-3756-9] [Medline: 28721628]
6. Bergstra SA, Murgia A, Te Velde AF, Caljouw SR. A systematic review into the effectiveness of hand exercise therapy in the treatment of rheumatoid arthritis. Clin Rheumatol 2014;33(11):1539-1548. [doi: 10.1007/s10067-014-2691-2] [Medline: 24952308]
7. Chaudhari P. The impact of rheumatoid arthritis and biologics on employers and payers. Biotechnol Healthc 2008;5(2):37-44 [FREE Full text] [Medline: 22478712]
8. Smolen JS, Aletaha D, McInnes IB. Rheumatoid arthritis. Lancet 2016;388(10055):2023-2038. [doi: 10.1016/S0140-6736(16)30173-8] [Medline: 27156434]
9. Verstappen SM. Rheumatoid arthritis and work: the impact of rheumatoid arthritis on absenteeism and presenteeism. Best Pract Res Clin Rheumatol 2015;29(3):495-511. [doi: 10.1016/j.berh.2015.06.001] [Medline: 26612244]
10. Metsios GS, Kitas GD. Physical activity, exercise and rheumatoid arthritis: effectiveness, mechanisms and implementation. Best Pract Res Clin Rheumatol 2018;32(5):669-682. [doi: 10.1016/j.berh.2019.03.013] [Medline: 31203925]
11. Williams MA, Srikesavan C, Heine PJ, Bruce J, Brosseau L, Hoxey-Thomas N, et al. Exercise for rheumatoid arthritis of the hand. Cochrane Database Syst Rev 2018;7(7):CD003832 [FREE Full text] [doi: 10.1002/14651858.CD003832.pub3] [Medline: 30063798]
12. Hammond A, Prior Y. The effectiveness of home hand exercise programmes in rheumatoid arthritis: a systematic review. Br Med Bull 2016;119(1):49-62. [doi: 10.1093/bmb/ldw024] [Medline: 27365455]
13. Porter BJ, Brittain A. Splinting and hand exercise for three common hand deformities in rheumatoid arthritis: a clinical perspective. Curr Opin Rheumatol 2012;24(2):215-221. [doi: 10.1097/BOR.0b013e3283503361] [Medline: 22269659]
14. O'Brien AV, Jones P, Mullis R, Mulherin D, Dziedzic K. Conservative hand therapy treatments in rheumatoid arthritis--a randomized controlled trial. Rheumatology (Oxford) 2006;45(5):577-583. [doi: 10.1093/rheumatology/kei215] [Medline: 16319099]
15. Lamb SE, Williamson EM, Heine PJ, Adams J, Dosanjh S, Dritsaki M, Strengthening and Stretching for Rheumatoid Arthritis of the Hand Trial (SARAH) Trial Team. Exercises to improve function of the rheumatoid hand (SARAH): a randomised controlled trial. Lancet 2015;385(9966):421-429. [doi: 10.1016/S0140-6736(14)60998-3] [Medline: 25308290]
16. Buljina AI, Taljanovic MS, Avdic DM, Hunter TB. Physical and exercise therapy for treatment of the rheumatoid hand. Arthritis Rheum 2001;45(4):392-397 [FREE Full text] [doi: 10.1002/1529-0131(200108)45:4<392::AID-ART353>3.0.CO;2-2] [Medline: 11501728]
17. Katz P, Andonian BJ, Huffman KM. Benefits and promotion of physical activity in rheumatoid arthritis. Curr Opin Rheumatol 2020;32(3):307-314. [doi: 10.1097/BOR.0000000000000696] [Medline: 32141951]
18. Conigliaro P, Triggianese P, De Martino E, Fonti GL, Chimenti MS, Sunzini F, et al. Challenges in the treatment of rheumatoid arthritis. Autoimmun Rev 2019;18(7):706-713. [doi: 10.1016/j.autrev.2019.05.007] [Medline: 31059844]
19. Manning VL, Hurley MV, Scott DL, Coker B, Choy E, Bearne LM. Education, self-management, and upper extremity exercise training in people with rheumatoid arthritis: a randomized controlled trial. Arthritis Care Res (Hoboken) 2014;66(2):217-227 [FREE Full text] [doi: 10.1002/acr.22102] [Medline: 23925924]
20. O'Brien L. The evidence on ways to improve patient's adherence in hand therapy. J Hand Ther 2012;25(3):247-250. [doi: 10.1016/j.jht.2012.03.006] [Medline: 22652329]
21. Cole T, Robinson L, Romero L, O'Brien L. Effectiveness of interventions to improve therapy adherence in people with upper limb conditions: a systematic review. J Hand Ther 2019;32(2):175-83.e2. [doi: 10.1016/j.jht.2017.11.040] [Medline: 29292028]
22. Roddy E, Zhang W, Doherty M, Arden NK, Barlow J, Birrell F, et al. Evidence-based recommendations for the role of exercise in the management of osteoarthritis of the hip or knee--the MOVE consensus. Rheumatology (Oxford) 2005;44(1):67-73. [doi: 10.1093/rheumatology/keh399] [Medline: 15353613]
23. Eccleston C, Blyth FM, Dear BF, Fisher EA, Keefe FJ, Lynch ME, et al. Managing patients with chronic pain during the COVID-19 outbreak: considerations for the rapid introduction of remotely supported (eHealth) pain management services. Pain 2020;161(5):889-893 [FREE Full text] [doi: 10.1097/j.pain.0000000000001885] [Medline: 32251203]
24. Berkovic D, Ackerman IN, Briggs AM, Ayton D. Tweets by people with arthritis during the COVID-19 pandemic: content and sentiment analysis. J Med Internet Res 2020;22(12):e24550 [FREE Full text] [doi: 10.2196/24550] [Medline: 33170802]
25. Slattery BW, Haugh S, O'Connor L, Francis K, Dwyer CP, O'Higgins S, et al. An evaluation of the effectiveness of the modalities used to deliver electronic health interventions for chronic pain: systematic review with network meta-analysis. J Med Internet Res 2019;21(7):e11086 [FREE Full text] [doi: 10.2196/11086] [Medline: 31317869]
26. Dias JF, Oliveira VC, Borges PR, Dutra FC, Mancini MC, Kirkwood RN, et al. Effectiveness of exercises by telerehabilitation on pain, physical function and quality of life in people with physical disabilities: a systematic review of randomised controlled

trials with GRADE recommendations. Br J Sports Med 2021;55(3):155-162. [doi: 10.1136/bjsports-2019-101375] [Medline: 33060156]
27. Pastora-Bernal JM, Martín-Valero R, Barón-López FJ, Estebanez-Pérez MJ. Evidence of benefit of telerehabitation after orthopedic surgery: a systematic review. J Med Internet Res 2017;19(4):e142 [FREE Full text] [doi: 10.2196/jmir.6836] [Medline: 28455277]
28. Blanquero J, Cortés-Vega MD, García-Frasquet M, Sánchez-Laulhé PR, Nieto Díaz de Los Bernardos MI, Suero-Pineda A. Exercises using a touchscreen tablet application improved functional ability more than an exercise program prescribed on paper in people after surgical carpal tunnel release: a randomised trial. J Physiother 2019;65(2):81-87 [FREE Full text] [doi: 10.1016/j.jphys.2019.02.008] [Medline: 30926400]
29. Blanquero J, Cortés-Vega MD, Rodríguez-Sánchez-Laulhé P, Corrales-Serra BP, Gómez-Patricio E, Díaz-Matas N, et al. Feedback-guided exercises performed on a tablet touchscreen improve return to work, function, strength and healthcare usage more than an exercise program prescribed on paper for people with wrist, hand or finger injuries: a randomised trial. J Physiother 2020;66(4):236-242 [FREE Full text] [doi: 10.1016/j.jphys.2020.09.012] [Medline: 33069608]
30. Dixon WG, Michaud K. Using technology to support clinical care and research in rheumatoid arthritis. Curr Opin Rheumatol 2018;30(3):276-281 [FREE Full text] [doi: 10.1097/BOR.0000000000000485] [Medline: 29369089]
31. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
32. Mollard E, Michaud K. A mobile app with optical imaging for the self-management of hand rheumatoid arthritis: pilot study. JMIR Mhealth Uhealth 2018;6(10):e12221 [FREE Full text] [doi: 10.2196/12221] [Medline: 30373732]
33. Seppen BF, Wiegel J, L'ami MJ, Duarte Dos Santos Rico S, Catarinella FS, Turkstra F, et al. Feasibility of self-monitoring rheumatoid arthritis with a smartphone app: results of two mixed-methods pilot studies. JMIR Form Res 2020;4(9):e20165 [FREE Full text] [doi: 10.2196/20165] [Medline: 32955447]
34. Voth EC, Oelke ND, Jung ME. A theory-based exercise app to enhance exercise adherence: a pilot study. JMIR Mhealth Uhealth 2016;4(2):e62 [FREE Full text] [doi: 10.2196/mhealth.4997] [Medline: 27307134]
35. Rodríguez-Sánchez-Laulhé P, Luque-Romero LG, Blanquero J, Suero-Pineda A, Biscarri-Carbonero Á, Barrero-García FJ, et al. A mobile app using therapeutic exercise and education for self-management in patients with hand rheumatoid arthritis: a randomized controlled trial protocol. Trials 2020;21(1):777 [FREE Full text] [doi: 10.1186/s13063-020-04713-4] [Medline: 32912305]
36. American College of Rheumatology Subcommittee on Rheumatoid Arthritis Guidelines. Guidelines for the management of rheumatoid arthritis: 2002 update. Arthritis Rheum 2002;46(2):328-346 [FREE Full text] [doi: 10.1002/art.10148] [Medline: 11840435]
37. Ellegaard K, von Bülow C, Røpke A, Bartholdy C, Hansen IS, Rifbjerg-Madsen S, et al. Hand exercise for women with rheumatoid arthritis and decreased hand function: an exploratory randomized controlled trial. Arthritis Res Ther 2019;21(1):158 [FREE Full text] [doi: 10.1186/s13075-019-1924-9] [Medline: 31242937]
38. Williams MA, Williamson EM, Heine PJ, Nichols V, Glover MJ, Dritsaki M, et al. Strengthening and stretching for rheumatoid arthritis of the hand (SARAH). A randomised controlled trial and economic evaluation. Health Technol Assess 2015;19(19):1-222 [FREE Full text] [doi: 10.3310/hta19190] [Medline: 25748549]
39. Rønningen A, Kjeken I. Effect of an intensive hand exercise programme in patients with rheumatoid arthritis. Scand J Occup Ther 2008;15(3):173-183. [doi: 10.1080/11038120802031129] [Medline: 19180723]
40. Williams MA, Heine PJ, Bruce J, Brosseau L, Lamb S. Exercise therapy for the rheumatoid hand. Cochrane Database Syst Rev 2012;4:CD003832. [doi: 10.1002/14651858.CD003832.pub2]
41. Ageberg E, Link A, Roos EM. Feasibility of neuromuscular training in patients with severe hip or knee OA: the individualized goal-based NEMEX-TJR training program. BMC Musculoskelet Disord 2010;11:126 [FREE Full text] [doi: 10.1186/1471-2474-11-126] [Medline: 20565735]
42. Sandal LF, Roos EM, Bøgesvang SJ, Thorlund JB. Pain trajectory and exercise-induced pain flares during 8 weeks of neuromuscular exercise in individuals with knee and hip pain. Osteoarthritis Cartilage 2016;24(4):589-592 [FREE Full text] [doi: 10.1016/j.joca.2015.11.002] [Medline: 26564575]
43. Thomeé R. A comprehensive treatment approach for patellofemoral pain syndrome in young women. Phys Ther 1997;77(12):1690-1703. [doi: 10.1093/ptj/77.12.1690] [Medline: 9413448]
44. Chung KC, Pillsbury MS, Walters MR, Hayward RA. Reliability and validity testing of the Michigan hand outcomes questionnaire. J Hand Surg Am 1998;23(4):575-587. [doi: 10.1016/S0363-5023(98)80042-7] [Medline: 9708370]
45. Barrios BI, Papasidero SB, Medina MA, Chaparro del Moral RE, Rillo OL, Paira SO, et al. Correlación de diferentes métodos de valoración de la rigidez matinal con índices de actividad y discapacidad en pacientes con artritis reumatoidea. Rev Argent Reumatol 2017;28(1):9-13. [doi: 10.47196/rar.v28i1.565]
46. Castarlenas E, de la Vega R, Jensen MP, Miró J. Self-report measures of hand pain intensity: current evidence and recommendations. Hand Clin 2016;32(1):11-19. [doi: 10.1016/j.hcl.2015.08.010] [Medline: 26611384]
47. ASHT's clinical assessment recommendations. 2nd edition. Chicago, IL: American Society of Hand Therapists; 1992:41-45.

48. Salaffi F, Di Carlo M, Carotti M, Farah S. Validity and interpretability of the QuickDASH in the assessment of hand disability in rheumatoid arthritis. Rheumatol Int 2019;39(5):923-932. [doi: 10.1007/s00296-018-4216-9] [Medline: 30511229]
49. Dritsaki M, Petrou S, Williams M, Lamb SE. An empirical evaluation of the SF-12, SF-6D, EQ-5D and Michigan hand outcome questionnaire in patients with rheumatoid arthritis of the hand. Health Qual Life Outcomes 2017;15(1):20 [FREE Full text] [doi: 10.1186/s12955-016-0584-6] [Medline: 28118833]
50. Waljee JF, Chung KC, Kim HM, Burns PB, Burke FD, Wilgis EF, et al. Validity and responsiveness of the Michigan hand questionnaire in patients with rheumatoid arthritis: a multicenter, international study. Arthritis Care Res (Hoboken) 2010;62(11):1569-1577 [FREE Full text] [doi: 10.1002/acr.20274] [Medline: 20521331]
51. Miranda D, Ramírez J, Rueda L, García J, Wolf G, Lugo A LH. Validación del "Michigan hand outcomes questionnaire" para población colombiana. Rev Colomb Reumatol 2008;15(4):271-290 [FREE Full text]
52. Sheehy C, Gaffney K, Mukhtyar C. Standardized grip strength as an outcome measure in early rheumatoid arthritis. Scand J Rheumatol 2013;42(4):289-293. [doi: 10.3109/03009742.2012.747624] [Medline: 23311733]
53. Higgins SC, Adams J, Hughes R. Measuring hand grip strength in rheumatoid arthritis. Rheumatol Int 2018;38(5):707-714. [doi: 10.1007/s00296-018-4024-2] [Medline: 29626222]
54. Kennedy D, Jerosch-Herold C, Hickson M. The reliability of one vs. three trials of pain-free grip strength in subjects with rheumatoid arthritis. J Hand Ther 2010;23(4):384-391. [doi: 10.1016/j.jht.2010.05.002] [Medline: 20971419]
55. MacDermid JC, Kramer JF, Woodbury MG, McFarlane RM, Roth JH. Interrater reliability of pinch and grip strength measurements in patients with cumulative trauma disorders. J Hand Ther 1994;7(1):10-14. [doi: 10.1016/s0894-1130(12)80035-4] [Medline: 8012479]
56. Sferra da Silva G, de Almeida Lourenço M, de Assis MR. Hand strength in patients with RA correlates strongly with function but not with activity of disease. Adv Rheumatol 2018;58(1):20. [doi: 10.1186/s42358-018-0020-1] [Medline: 30657069]
57. Hammond A, Prior Y, Tyson S. Linguistic validation, validity and reliability of the British English versions of the Disabilities of the Arm, Shoulder and Hand (DASH) questionnaire and QuickDASH in people with rheumatoid arthritis. BMC Musculoskelet Disord 2018;19(1):118 [FREE Full text] [doi: 10.1186/s12891-018-2032-8] [Medline: 29661183]
58. Shauver MJ, Chung KC. The minimal clinically important difference of the Michigan hand outcomes questionnaire. J Hand Surg Am 2009;34(3):509-514 [FREE Full text] [doi: 10.1016/j.jhsa.2008.11.001] [Medline: 19258150]
59. Sobue Y, Kojima T, Ito H, Nishida K, Matsushita I, Kaneko Y, et al. Does exercise therapy improve patient-reported outcomes in rheumatoid arthritis? A systematic review and meta-analysis for the update of the 2020 JCR guidelines for the management of rheumatoid arthritis. Mod Rheumatol 2022;32(1):96-104 [FREE Full text] [doi: 10.1080/14397595.2021.1886653] [Medline: 33565350]
60. Williamson E, McConkey C, Heine P, Dosanjh S, Williams M, Lamb SE. Hand exercises for patients with rheumatoid arthritis: an extended follow-up of the SARAH randomised controlled trial. BMJ Open 2017;7(4):e013121 [FREE Full text] [doi: 10.1136/bmjopen-2016-013121] [Medline: 28404610]
61. Dellhag B, Wollersjö I, Bjelle A. Effect of active hand exercise and wax bath treatment in rheumatoid arthritis patients. Arthritis Care Res 1992;5(2):87-92. [doi: 10.1002/art.1790050207] [Medline: 1390969]
62. Srikesavan C, Williamson E, Cranston T, Hunter J, Adams J, Lamb SE. An online hand exercise intervention for adults with rheumatoid arthritis (mySARAH): design, development, and usability testing. J Med Internet Res 2018;20(6):e10457 [FREE Full text] [doi: 10.2196/10457] [Medline: 29950288]
63. Srikesavan C, Williamson E, Thompson JY, Cranston T, Swales C, Lamb SE. The online version of an evidence-based hand exercise program for people with rheumatoid arthritis: a mixed-method, proof-of-concept study. J Hand Ther. Preprint posted online on October 26, 2020. [doi: 10.1016/j.jht.2020.10.011] [Medline: 33279362]
64. Luo D, Wang P, Lu F, Elias J, Sparks JA, Lee YC. Mobile apps for individuals with rheumatoid arthritis: a systematic review. J Clin Rheumatol 2019;25(3):133-141. [doi: 10.1097/RHU.0000000000000800] [Medline: 29933327]
65. Tonga E, Williamson E, Srikesavan C, Özen T, Sarıtaş F, Lamb SE. A hand exercise mobile app for people with rheumatoid arthritis in Turkey: design, development and usability study. Rheumatol Int 2021;41(6):1151-1160. [doi: 10.1007/s00296-021-04860-0] [Medline: 33870452]
66. Martinez-Calderon J, Meeus M, Struyf F, Luque-Suarez A. The role of self-efficacy in pain intensity, function, psychological factors, health behaviors, and quality of life in people with rheumatoid arthritis: a systematic review. Physiother Theory Pract 2020;36(1):21-37. [doi: 10.1080/09593985.2018.1482512] [Medline: 29873569]
67. Rubio Oyarzun D, Neuber Slater K, Araya Quintanilla F, Gutierrez Espinoza H, Arias Poblete L, Olguin Huerta C, et al. Intervención de ejercicios de habilidad motora manual en el dolor y función en sujetos adultos con artritis reumatoide: serie de casos. Rev Soc Esp Dolor 2017;24(6):294-303. [doi: 10.20986/resed.2018.3576/2017]
68. Brorsson S, Hilliges M, Sollerman C, Nilsdotter A. A six-week hand exercise programme improves strength and hand function in patients with rheumatoid arthritis. J Rehabil Med 2009;41(5):338-342 [FREE Full text] [doi: 10.2340/16501977-0334] [Medline: 19363566]
69. Dogu B, Sirzai H, Yilmaz F, Polat B, Kuran B. Effects of isotonic and isometric hand exercises on pain, hand functions, dexterity and quality of life in women with rheumatoid arthritis. Rheumatol Int 2013;33(10):2625-2630. [doi: 10.1007/s00296-013-2787-z] [Medline: 23739921]

70. Wessel J. The effectiveness of hand exercises for persons with rheumatoid arthritis: a systematic review. J Hand Ther 2004;17(2):174-180. [doi: 10.1197/j.jht.2004.02.006] [Medline: 15162104]
71. Martinez-Calderon J, Zamora-Campos C, Navarro-Ledesma S, Luque-Suarez A. The role of self-efficacy on the prognosis of chronic musculoskeletal pain: a systematic review. J Pain 2018;19(1):10-34. [doi: 10.1016/j.jpain.2017.08.008] [Medline: 28939015]
72. Brighton SW, Lubbe JE, van der Merwe CA. The effect of a long-term exercise programme on the rheumatoid hand. Br J Rheumatol 1993;32(5):392-395. [doi: 10.1093/rheumatology/32.5.392] [Medline: 8495260]
73. Vliet Vlieland TP, Zwinderman AH, Breedveld FC, Hazes JM. Measurement of morning stiffness in rheumatoid arthritis clinical trials. J Clin Epidemiol 1997;50(7):757-763. [doi: 10.1016/s0895-4356(97)00051-6] [Medline: 9253385]
74. Haugen IK, Aaserud J, Kvien TK. Get a grip on factors related to grip strength in persons with hand osteoarthritis: results from an observational cohort study. Arthritis Care Res (Hoboken) 2021;73(6):794-800. [doi: 10.1002/acr.24385] [Medline: 32702166]
75. Bobos P, Nazari G, Szekeres M, Lalone EA, Ferreira L, MacDermid JC. The effectiveness of joint-protection programs on pain, hand function, and grip strength levels in patients with hand arthritis: a systematic review and meta-analysis. J Hand Ther 2019;32(2):194-211. [doi: 10.1016/j.jht.2018.09.012] [Medline: 30587434]
76. Hu H, Xu A, Gao C, Wang Z, Wu X. The effect of physical exercise on rheumatoid arthritis: an overview of systematic reviews and meta-analysis. J Adv Nurs 2021;77(2):506-522. [doi: 10.1111/jan.14574] [Medline: 33176012]
77. Aktekin LA, Eser F, Başkan BM, Sivas F, Malhan S, Öksüz E, et al. Disability of arm shoulder and hand questionnaire in rheumatoid arthritis patients: relationship with disease activity, HAQ, SF-36. Rheumatol Int 2011;31(6):823-826. [doi: 10.1007/s00296-010-1568-1] [Medline: 20680284]
78. Bearne LM, Coomer AF, Hurley MV. Upper limb sensorimotor function and functional performance in patients with rheumatoid arthritis. Disabil Rehabil 2007;29(13):1035-1039. [doi: 10.1080/09638280600929128] [Medline: 17612988]
79. Prodinger B, Hammond A, Tennant A, Prior Y, Tyson S. Revisiting the disabilities of the arm, shoulder and hand (DASH) and QuickDASH in rheumatoid arthritis. BMC Musculoskelet Disord 2019;20(1):41 [FREE Full text] [doi: 10.1186/s12891-019-2414-6] [Medline: 30683082]
80. Pedersini P, Negrini S, Cantero-Tellez R, Bishop MD, Villafañe JH. Pressure algometry and palpation of the upper limb peripheral nervous system in subjects with hand osteoarthritis are repeatable and suggest central changes. J Hand Ther 2020;33(1):103-111. [doi: 10.1016/j.jht.2018.06.003] [Medline: 30679091]
81. Rice D, McNair P, Huysmans E, Letzen J, Finan P. Best evidence rehabilitation for chronic pain part 5: osteoarthritis. J Clin Med 2019;8(11):1769 [FREE Full text] [doi: 10.3390/jcm8111769] [Medline: 31652929]
82. van Riel P, Alten R, Combe B, Abdulganieva D, Bousquet P, Courtenay M, et al. Improving inflammatory arthritis management through tighter monitoring of patients and the use of innovative electronic tools. RMD Open 2016;2(2):e000302 [FREE Full text] [doi: 10.1136/rmdopen-2016-000302] [Medline: 27933206]
83. de Thurah A, Stengaard-Pedersen K, Axelsen M, Fredberg U, Schougaard LM, Hjollund NH, et al. Tele-health followup strategy for tight control of disease activity in rheumatoid arthritis: results of a randomized controlled trial. Arthritis Care Res (Hoboken) 2018;70(3):353-360. [doi: 10.1002/acr.23280] [Medline: 28511288]

Abbreviations

ADL: activities of daily living
DMARD: disease-modifying antirheumatic drug
EXTRA: Education, Self-Management, and Upper Extremity Exercise Training
GS: grip strength
MHQ: Michigan Hand Outcome Questionnaire
QuickDASH: shortened version of the Disabilities of the Arm, Shoulder, and Hand questionnaire
RA: rheumatoid arthritis
SARAH: Strengthening and Stretching for Rheumatoid Arthritis of the Hand
VAS: visual analog scale

Edited by L Buis; submitted 05.12.21; peer-reviewed by H Mehdizadeh, H Ayatollahi; comments to author 29.12.21; revised version received 01.02.22; accepted 18.02.22; published 07.04.22.

Please cite as:
Rodríguez Sánchez-Laulhé P, Luque-Romero LG, Barrero-García FJ, Biscarri-Carbonero Á, Blanquero J, Suero-Pineda A, Heredia-Rizo AM
An Exercise and Educational and Self-management Program Delivered With a Smartphone App (CareHand) in Adults With Rheumatoid Arthritis of the Hands: Randomized Controlled Trial
JMIR Mhealth Uhealth 2022;10(4):e35462
URL: https://mhealth.jmir.org/2022/4/e35462
doi:10.2196/35462
PMID:35389367

Original Paper

Problem-Based mHealth Literacy Scale (PB-mHLS): Development and Validation

Lingmin Zhang[1], PhD; Pengxiang Li[2], PhD

[1]School of Journalism and Communication, Guangzhou University, Guangzhou, China

[2]School of Media and Communication, Shenzhen University, Shenzhen, China

Corresponding Author:
Pengxiang Li, PhD
School of Media and Communication
Shenzhen University
#L7-910, Houhai AVE
Nanshan District
Shenzhen, 518060
China
Phone: 86 0755 86581227
Email: lipx@szu.edu.cn

Abstract

Background: Mobile devices have greatly facilitated the use of digital health resources, particularly during the COVID-19 pandemic. Mobile health (mHealth) has become a common and important way to monitor and improve health conditions for people from different social classes. The ability to utilize mHealth affects its effectiveness; therefore, the widespread application of mHealth technologies calls for an instrument that can accurately measure health literacy in the era of mobile media.

Objective: We aimed to (1) identify the components of mHealth literacy for ordinary users and (2) develop a systematic scale for appropriately measuring individuals' self-perceived mHealth literacy through a problem-based framework.

Methods: We conducted an exploratory study involving in-depth interviews and observations (15 participants) in January 2020 and used exploratory factor analysis and confirmatory factor analysis to identify the components of mHealth literacy and develop an item pool. In February 2020, we conducted a pilot survey with 148 participants to explore the factor structures of items identified during the exploratory study. Subsequently, 2 surveys were administrated using quota sampling. The first survey (conducted in Guangdong, China) collected 552 responses during March 2020; we assessed composite reliability, convergent validity, and discriminant validity. The second survey (conducted in China nationwide) collected 433 responses during October 2021; we assessed criterion-related validity using structural equation modeling.

Results: We identified 78 items during the exploratory study. The final scale—the Problem-Based mHealth Literacy Scale—consists of 33 items that reflect 8 domains of mHealth literacy. The first web-based survey suggested that mHealth literacy consists of 8 factors (ie, subscales), namely, mHealth desire, mobile phone operational skills, acquiring mHealth information, acquiring mHealth services, understanding of medical terms, mobile-based patient–doctor communication, evaluating mHealth information, and mHealth decision-making. These factors were found to be reliable (composite reliability >0.7), with good convergent validity (average variance extracted >0.5) and discriminant validity (square root of average variance extracted are greater than the correlation coefficients between factors). The findings also revealed that these 8 factors should be grouped under a second-order factor model (χ^2/df=2.701; comparative fit index 0.921; root mean square error of approximation 0.056; target coefficient 0.831). The second survey revealed that mHealth use had a significant impact (β=0.43, P<.001) on mHealth literacy and that mHealth literacy had a significant impact (β=0.23, P<.001) on health prevention behavior.

Conclusions: This study revealed the distinctiveness of mHealth literacy by placing mHealth needs, the ability to understand medical terms, and the skills in patient–doctor interactions in the foreground. The Problem-Based mHealth Literacy Scale is a useful instrument for comprehensively measuring individuals' mHealth literacy and extends the concept of health literacy to the context of mobile communication.

(JMIR Mhealth Uhealth 2022;10(4):e31459) doi:10.2196/31459

KEYWORDS

mobile health; mHealth literacy; instrument development; problem-based framework

Introduction

Background

Mobile technologies afford users ubiquitous access to information from the internet and to digital apps. Nearly half of the world's current population (48.5%) owns a smart mobile phone, which has become the dominant form of access to the internet [1]. In countries such as China, this rate is even higher—99.6% of Chinese internet users rely on their mobile phones to access the internet [2]. The mobility, multimodality, and interactivity of mobile phones help individuals easily access digital health resources to manage and improve their health conditions. People can conveniently use health apps to monitor their health conditions, facilitate their physical exercises, acquire health information, and consult doctors. During the COVID-19 pandemic, mobile health (mHealth) has become particularly important. People rely on their mobile phones to check the status of the pandemic in their countries or neighborhoods, search for strategies for health prevention and protection, register for vaccines, and for many other activities. Accordingly, the adoption of mobile technologies for health-related practices has received widespread attention from researchers, practitioners, and policy makers [3,4]. As an extension of eHealth, mHealth generally refers to the use of mobile and wireless technologies, especially mobile phones and tablets, for health information and improving health care services and health outcomes [5].

Previous studies [6-8] have shown that digital health resources can benefit patients with chronic diseases by making self-monitoring more convenient, simplifying administrative procedures in health care, increasing access to health care services, reducing the cost of health care, and enriching resources in low- and middle-income countries. However, an individual's health can be negatively impacted if they have poor mHealth skills or if they fail to adopt digital technology [9]. Biased digital health information, as well as the unskilled use of digital technology, can lead to misdiagnosis, which in turn may lead patients to try unapproved or unreliable therapies and to miss opportunities for optimal care [10,11]. Since mobile devices have become the main channel by which web-based health resources are accessed, a person's mHealth literacy has become a factor in their health.

The concept of mHealth literacy builds upon that of eHealth literacy and is applied in the context of mobile and wireless technologies. eHealth literacy refers to "the ability to seek, find, understand, and appraise health information from electronic sources, and apply the knowledge gained to addressing or solving a health problem [12]." mHealth literacy is generally defined as the ability to use health-related apps on a mobile phone [13] or the ability to use mobile devices to search, find, understand, appraise, and apply health information to address a health problem [14]. Although scales for measuring eHealth literacy (eHEALS 1.0 [12] and 2.0 [15]) have been developed, there is no appropriate scale for assessing the unique features of mHealth literacy.

It is of both theoretical and practical importance to develop a scale for measuring mHealth literacy. Theoretically, the notion of health literacy continuously changes in accordance with emerging information technologies. Conventional health literacy concentrates on the abilities to comprehend, evaluate, and apply health information. However, when assessing eHealth literacy, the skills to acquire health information (eg, computer operation skills) and interact with others for support (eg, social interaction) via digital devices are also considered to be crucial components. Likewise, the skill sets for approaching and acquiring health-related resources via mobile phone are becoming more complex. For instance, eHealth literacy scales mainly examine ability to comprehend and evaluate health information, whereas mHealth literacy encompasses not only health information but also important digital resources (eg, health apps, real-time web-based doctor consulting) for health management and improvement [13]. Thus, the usability and multimodality of mobile media allow individuals to conveniently access to various health-related resources. The resources and required skills for mHealth are distinct; yet, they have not been comprehensively represented in existing scales.

Practically, eHealth literacy scales examine the abilities to acquire health information via desktops or laptops, which favors educated and affluent populations; eHealth may not be available to some people, especially those with limited access to computers. In comparison, the convenience and accessibility of mobile devices have led these devices to become pervasive globally; in turn, this has lowered barriers to the use of web-based resources. For instance, more than half of mobile phone users (59.6%) in China are not educated above the junior high level [2]. Therefore, it is essential to design an mHealth literacy scale for diverse population groups, in order to characterize their abilities to access digital health resources.

Previous studies [14,16] have measured mHealth literacy by altering certain phrases in existing eHealth literacy scales (eg, changing "on the internet" to "on a mobile phone"), but these simple word-level modifications do not adequately capture the distinctiveness of mHealth. As reflected in the transition from conventional health literacy scales to eHealth literacy scales, each new emerging health literacy scale is built on the established scales but also shows unique characteristics. Although the concept of mHealth literacy is an extension of eHealth literacy and the domains of these two concepts overlap to some extent, the assessment of mHealth literacy is fundamentally different from that of eHealth literacy. Therefore, mHealth literacy assessment tools need to account for the distinct features of mHealth.

Objective

We aimed to identify the fundamental components of mHealth literacy for ordinary users and develop a systematic self-report scale that can be used to appropriately measure individuals perceptions of their own mHealth literacy through a problem-based framework.

Review of eHealth Scales

To operationalize mHealth literacy, it is essential to first understand how health literacy, especially eHealth literacy, is measured in the literature. Conventional health literacy scales (eg, The Test of Functional Health Literacy in Adults [17]) focus on individuals' abilities to read and comprehend health information as well as their ability to accurately express health problems. eHealth literacy scales emphasize individuals' abilities to access digital resources to improve health conditions.

There are 3 widely used and empirically validated instruments for assessing eHealth literacy—2 versions of the Electronic Health Literacy Scale, and the Digital Health Literacy Instrument. eHEALS has been widely adopted since the era of Web 1.0 [18,19]. eHEALS comprises 8 self-reported items that measure an individual's ability to acquire, appraise, and apply health information [18]. However, this scale ignores the fact that people can generate health-related information and interact with each other on the internet. Accordingly, eHEALS 2.0, which incorporates a specific dimension to measure social media interactions, can assess health literacy in the era of Web 2.0 [15]. Although eHEALS 1.0 and 2.0 are valid and reliable instruments, they examine individuals' abilities to acquire, appraise, produce, and apply health-related resources at a relatively general level—each specific domain (ie, acquiring, appraising, producing, and applying web-based health resources) is measured by a single item. The Digital Health Literacy Instrument [16] addresses this problem by measuring 7 domains of health literacy using 21 self-reported items—each domain is measured by at least three items.

In addition, the Patient Readiness to Engage in Health Internet Technology provides a useful conceptual framework for measuring patient literacy in processing web-based health information. This framework [20] includes eight domains: (1) health information needs; (2) computer or internet experience; (3) computer anxiety; (4) the preferred mode of interaction; (5) the relationship with the doctor; (6) mobile phone expertise; (7) internet privacy; and (8) "no news is good news." Despite not being validated with empirical data, this conceptual framework still provides insight into the components embedded in digital health literacy.

Because mHealth is an extension of eHealth to the context of mobile communication, eHealth tools can provide the foundation for developing a new mHealth literacy instrument. Indeed, 5 basic domains can be extracted from existing eHealth literacy scales. These include the ability to use digital devices, and the abilities to acquire, comprehend, appraise, and apply health resources with digital devices. In addition, 2 domains proposed under the Patient Readiness to Engage in Health Internet Technology framework—the need for health information and computer anxiety (and internet privacy)—are informative for mHealth literacy because they reflect the motivations and barriers to using electronic and digital health resources. These 7 domains can serve as the foundation upon which the indicators of mHealth literacy are developed.

The Problem-Based Approach

Most eHealth literacy scales use self-reported items. That is, these scales focus on individuals' subjective perceptions of and experiences using digital health resources, given that it is difficult to comprehensively measure the operational skills of digital devices (be it a computer or a mobile phone) through an objective test. Nonetheless, such skills are crucial components in eHealth literacy. Self-reported measurement has a number of advantages, namely, easy interpretability, richness of information, motivation to report, causal force, and sheer practicality [21,22] and is thus applicable and useful. Self-perceptions strongly influence the ways in which individuals interact with the world, and subjective experience form the basis of health literacy enhancement interventions [23,24]. In line with previous research [12,14-16,18,19], in this study, we considered the concept of mHealth literacy to be individuals' perceived abilities in utilizing and managing mobile health resources.

A problem-based approach is particularly suitable for exploring the distinctive features of mobile health resources and clarifying the structure of mHealth literacy. The problem-based framework emphasizes a person's subjective experiences in solving specific issues. The problem-based framework is modeled after the problem-based learning framework used in medical education, which was initially introduced as a student-centered pedagogy, in which students spontaneously and autonomously learned by solving specific problems [25]. By analyzing a given problem, students were able to determine which skills and attributes they required, and further developed lifelong learning abilities [26].

Methods

Overview

Given that an mHealth literacy instrument should evaluate an individual's ability to solve daily health-related issues, it is reasonable to adopt the problem-based approach as the underlying framework for such an instrument. This study comprehensively examines individuals' subjective experiences of daily health care practices, including their perception and comprehension of health problems, and their abilities to solve such problems. Specifically, drawing from the problem-based framework, we employ observations and in-depth interviews to explore the specific situations in which mobile users will exploit mHealth resources, as well as to understand the distinct behavioral trajectories of mHealth. Subsequently, we integrate the exploratory findings and the indicators extracted from existing scales to develop the specific indicators.

We conducted in-depth interviews and also observed users to determine their daily mHealth practices, and to identify the specific domains. Then, we assembled a pool of candidate items, based on the domains that we identified, and pretested the candidate items and scale with a small sample. We used item analysis and exploratory factor analysis to identify the most informative items for each domain. Finally, we used confirmatory factor analysis and structural equation modeling to validate and refine the scale that we developed in 2 surveys with large sample sizes.

Ethics

All empirical studies, as part of a large research project, have been approved by the South China University of Technology (IRB00013151). Participants were informed both the purpose and the process in advance. They were voluntarily involved in the project and could withdraw at any time.

Participants

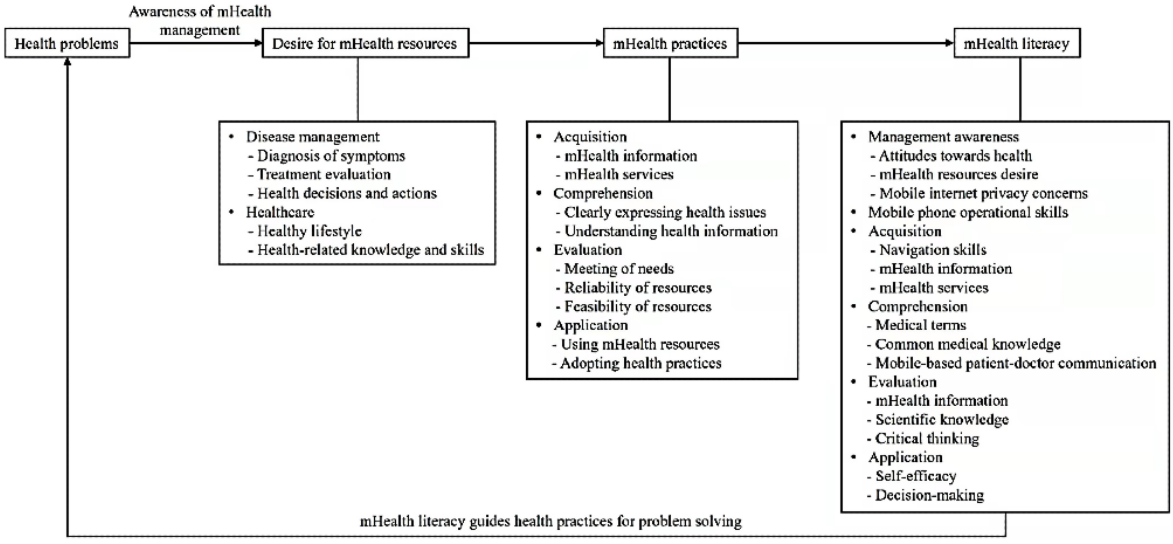
All mobile phone users older than 18 years were eligible to participate.

Exploratory Study to Identify the Domains of mHealth Literacy

In January 2020 in Guangzhou, China, we invited 15 users to investigate how they used their mobile phones to address health problems. Each participant was randomly assigned 1 of 3 types of hypothetical health problems. Participants were asked to read the instructions and then solve the problems with their mobile phones (Multimedia Appendix 1). During the problem-solving process, we observed and recorded participants' behavioral trajectories of mobile phone use and conducted in-depth interviews to understand their cognitive processes when performing specific actions.

We initially identified six main types of mHealth literacy skills (Figure 1): (1) the awareness of one's health management (including one's health attitudes, desire for mHealth, and concerns about internet security); (2) operational skills for individuals; (3) the ability to acquire mHealth resources (including navigation skills, information searching skills, and searching for mHealth services via mobile devices); (4) the ability to understand medical information contained in mHealth resources (including the knowledge of medical terms, common medical knowledge, and mobile-based patient–doctor communication skills); (5) the ability to evaluate mHealth resources (including the ability to assess the credibility of mHealth information, as well as the levels of the scientific knowledge and critical thinking skills); and (6) the ability to apply mHealth resources (including self-efficacy, and the ability to make decisions based on mHealth). Since each main type of health literacy skill contained a few key skills, we distributed them across 15 domains in a problem-based framework to guide the construction of our mHealth literacy instrument.

Figure 1. Problem-based framework for mHealth practices and literacy.



Development of the Scale

We assembled an item pool by modifying items from existing eHealth literacy scales and creating new items based on the domains identified in the exploratory study. We developed a minimum of 5 items for each domain to ensure that sufficient candidate items were selected. We invited 25 mobile users to test the readability of these items—to check whether their descriptions were comprehensible and consistent with common practices. Based on their feedback, we modified, merged, and deleted various items. The final pool contained 78 items; we assigned a 5-point Likert scale to each of the 78 items.

We then conducted a pilot survey to analyze the psychometrics of the items, determine the reliability of the subscales, and select the most discriminatory items. After recoding reverse items, we performed item discrimination analysis, corrected item-total correlations, and exploratory factor analysis to identify the items that best discriminated the domains and subscales.

We arranged participants' scores in ascending order and defined those who ranked in the top 27% as high scoring, and those who ranked in the lowest 27% as low scoring. The pass rate was defined as the percentage of participants in a group with 4 or 5 as the response on the 5-point Likert scale; therefore, the *discrimination index = pass rate for high-scoring participants − pass rate for low-scoring participants*. In this study, we excluded items with a discrimination index <0.3.

Corrected item-total correlations between an item and its domain were calculated; we eliminated items with values less than 0.5 [27].

We used exploratory factor analysis to explore the structure of the items. Studies suggest that approximately 100 to 200 participants should be surveyed in order to have sufficient data for conducting an exploratory factor analysis [28,29]. Thus, we conducted a web-based survey of 148 adult mobile users in February 2020 in Guangdong province, China. We used quota

sampling, based on age and sex, to ensure that the demographic characteristics of the sample were representative of the general population of mobile users in Guangdong. We used a principal component analysis with a varimax rotation to extract the factors.

Notably, we did not rely entirely on statistical criteria to select items and construct domains. This was because the selection of variables should be guided by theory and the findings of previous studies [30]. Therefore, we modified the factors and items considering the theoretical frame and statistical results.

Validation of the Scale

Overview

We conducted 2 web-based surveys with large samples. The survey studies were administrated with the Survey Plus applet and Tencent Questionnaire website. Studies indicate that the ratio of items to participants should be between 5 and 10 when confirmatory factor analysis is performed to validate a scale [31]. As with the pilot survey, we used quota sampling (based on gender and age) in both surveys to ensure that the samples sufficiently captured specific demographic characteristics of general internet users.

First Survey

The first survey was administrated in March 2020 in Guangdong province, China. We recruited 552 mobile users through a web-based survey to test the scale. The first survey was used to validate the scale that was identified in the pilot test. We first ran confirmatory factor analysis separately for each of the 10 factors and subscales to examine the internal reliability, construct validity, and discriminant validity of each factor and identify a model that fit the empirical data accurately.

Specifically, we evaluated the confirmatory factor analysis model using the *P* value and other indices, including the χ^2/df ratio, which is acceptable if the value is between 2 and 5; the goodness of fit index (GFI), which is acceptable if the value is >0.8; the adjusted GFI, which is acceptable if the value is >0.8; the comparative fit index, which is acceptable if the value is >0.9; the nonnormed fit index, which is acceptable if the value is>0.9; and the root mean square error of approximation (RMSEA), which is acceptable if the value is <0.08, with a value <0.05 being optimal [30]. To test the convergent validity of each factor, we evaluated the composite reliability (acceptable if the value is >0.7), the average variance extracted (acceptable if the value is >0.5) [32], and the factor loading of each item (acceptable if the value is >0.6) [33]. We deleted items that had factor loadings less than 0.6 and items with large residual correlations.

Then we compared models by checking goodness of fit indices and their target coefficients. The target coefficient is an index that examines the appropriateness of using a second-order model to substitute a first-order model. This coefficient ranges from 0 to 1, as it calculates the ratio of the chi-square values of the first-order model to those of the second-order model, with acceptable values above 0.7. The closer the target coefficient is to 1, the less information the second-order model loses, and the more suitable it is to use the second-order model in place of the first-order model [34].

Second Survey

The second survey, in October 2021, was used to validate the internal and external reliability and the criterion-related validity of the scale. We recruited valid participants nationwide (in China) to improve the generalizability of the results. We adopted exploratory factor analysis and confirmatory factor analysis to retest the validity of the 8 factors and the whole second-order model.

To verify the appropriateness of applying the model to different groups and populations, we conducted multiple-group analysis, using educational level and age, respectively. In addition, to examine the criterion-related validity, we introduced 2 factors that were likely to be associated with health literacy: mHealth use and health prevention behaviors.

mHealth use was assessed with the multiple-choice question "In the past three months, which of the following mHealth behaviors have you engaged in?" with 11 response options (eg, "searching for health or disease information on a mobile phone" and "using health apps on a mobile phone"). The more options the participants selected, the higher the score of mHealth use they received.

Health prevention behaviors (Cronbach α=0.895) was assessed with a 5-point Likert scale. The 10 items in this scale were adapted from the literature [35], where they were mainly used for assessing preventive and protective behaviors during the COVID-19 pandemic (eg, "washing hands after arriving home" and "covering the mouth and nose with a tissue or sleeves when coughing or sneezing").

Exploratory factor analysis was conducted using SPSS (version 26.0; IBM Corp). Confirmatory factor analysis and structural equation modeling were conducted using AMOS (version 22.0; IBM Corp).

Results

Pilot Study

On the basis of our exploratory study, we assembled a pool of 78 items reflecting 15 domains of mHealth literacy. In the pilot test sample, 52% (77/148) of the participants were men and 48% (71/148) were women, and the majority of participants were between 30 and 60 years old (83/148, 56%). After item discrimination analysis and corrected item-total correlations calculation, 60 items remained. We then performed an exploratory factor analysis to extract factors that could reveal the structure of these items.

The Kaiser-Meyer-Olkin value was 0.893, and the Bartlett test of sphericity was statistically significant (*P*<.001), which indicated that exploratory factor analysis was appropriate. We extracted 12 common factors that had a cumulative variance contribution rate of 77.0%. The domains *self-efficacy* and *health practice* were merged; the other factors remained unchanged. After eliminating cross-loaded items and low factor-loading items, 45 items remained. Exploratory factor analysis was repeated a (Kaiser-Meyer-Olkin 0.898; Bartlett test of sphericity

P<.001), and 11 common factors were extracted from the remaining 45 items, with a cumulative variance contribution rate of 78.7% (Multimedia Appendix 2). The results also indicated a good internal consistency for each factor (all values of Cronbach's alpha are between 0.777 and 0.921).

We modified the working scale slightly to reflect theory concerning mHealth. The modifications included (1) placing the item "I can identify relevant health information from search results" under Factor 4 (searching for information on a mobile phone); (2) moving the item "The health information accessed from a mobile phone is reliable" to Factor 10 (critical appraisal of information); (3) combining Factor 3 (mobile phone navigation skills) and Factor 4 (searching for information on a mobile phone) to form a new factor, that is, the acquisition of mHealth information; and (4) the addition of a new item, "I can tell whether the acquired health information can solve my problems" into Factor 6 (understanding of medical terms). In sum, we developed a measurement scale with 46 items under 10 subscales scored with a 5-point Likert scale.

First Validation Survey

A total of 552 respondents (18 to 24 years: n=70, 12.7%; 25 to 30 years: n=134, 24.3%; 31 to 40 years: n=147, 26.6%; 41 to 50 years: n=148, 26.8%; older than 50 years: n=53, 9.6%) participated in the first validation survey. Of these respondents, 52.2% (n=288) were male and 47.8% (n=264) were women; 42.6% (n=235) of participants possessed a high school education (or a lower educational level), the average family monthly income was between 3000 RMB (1 RMB is equivalent to approximately US $0.16) and 10,000 RMB (49.0%, n=271), and 71.0% (n=391) lived in urban areas.

We first ran confirmatory factor analysis separately for each of the 10 factors. The results suggested that the factors were unacceptable, namely, science knowledge and information critique. Therefore, we deleted the factor *science knowledge*, and integrated the acceptable item (information critique) into the factor *evaluating mobile health information*. After deleting items with low factor loadings and items with large residual correlations, 8 factors were retained. The factor loading of each item was larger than 0.6; composite reliability values were between 0.78 and 0.88, and the average variance extracted value of each factor was greater than 0.5 (Table 1).

The square root of average variance extracted for each factor and the Pearson correlation coefficients between factors revealed good construct validity and discriminant validity among the 8 factors (Table 2).

We compared 3 models to determine which model fit the data best (Figure 2). Model 1 was a single-factor model, in which each item corresponded to an overarching factor. Model 2 was a first-order factor model that included 8 factors that correlated with each other (ie, each item corresponded to a certain factor and subscale, and these 8 factors were correlated with one another). Model 3 was an 8-factor second-order model (ie, each item corresponded to 1 of the 8 factors in the first order, and these 8 factors corresponded to an overarching factor in the second order).

Model 1 could not adequately fit the data; its indices of goodness of model fit fell out of the acceptable range (Table 3); however, the indices for models 2 and 3 were within the acceptable ranges. The target coefficient for both model 2 and model 3 was 0.831 (ie, 1092.865/1315.368). Thus, although both models fit the data well, Model 3 (ie, the second-order model with 8 factors) better reflected the structure of the data. Consequently, we used a second-order model composed of 33 items in 8 domains and subscales to build the Problem-Based mHealth Literacy Scale (PB-mHLS).

Table 1. Confirmatory factor analysis results.

Factors and items	Standard Factor Loading	Squared multiple correlations	Composite reliability	Average variance extracted
Desire for mHealth When encountering health problems that I do not know how to deal with help me.	—[a]	—	0.78	0.54
S101: I search the mobile internet for health information.	0.637	0.41	—	—
S102: The information found on the mobile internet can	0.821	0.67	—	—
S103: I feel that using the mobile internet is a convenient way to solve the problem.	0.731	0.53	—	—
Mobile phone operational skills	—	—	0.88	0.59
S201: I can operate mobile phones easily.	0.656	0.44	—	—
S203: I know how to download new apps.	0.822	0.67	—	—
S204: I know how to enter keywords into a search box.	0.764	0.58	—	—
S205: I know how to follow officially authenticated accounts on social media.	0.821	0.67	—	—
S206: I can successfully purchase goods web-based with my mobile phone.	0.765	0.59	—	—
Acquiring mHealth information	—	—	0.88	0.66
S301: I know what health resources are available on the mobile internet.	0.825	0.75	—	—
S302: I know where to find helpful health resources on the mobile internet.	0.855	0.73	—	—
S303: I know how to find mobile-based health resources using my mobile phone.	0.863	0.68	—	—
S305: I know how to enter keywords into a search box to find the health resources I need.	0.681	0.46	—	—
Acquiring mHealth services	—	—	0.87	0.68
S308: I can make a doctor's appointment using my mobile phone.	0.728	0.53	—	—
S309: I know that it is possible to see a doctor for a one-to-one consultation using my mobile phone.	0.848	0.72	—	—
S310: I can complete a mobile-based medical consultation with a doctor using my mobile phone.	0.895	0.80	—	—
Understanding of medical terms	—	—	0.83	0.62
S402: I can understand the explanations given when searching for information about certain symptoms on the mobile internet.	0.757	0.57	—	—
S403: I can evaluate the severity of a disease according to the description given on the mobile internet.	0.784	0.62	—	—
S404: I can tell whether the health information can solve my problems.	0.823	0.68	—	—
Mobile-based patient–doctor communication	—	—	0.86	0.55
S405: I can clearly describe my health conditions to an web-based doctor during a mobile phone-based consultation.	0.749	0.56	—	—
S406: I can tell the doctor which medicines I am taking during a mobile phone-based consultation.	0.764	0.58	—	—
S407: I know that it is possible to take photos of relevant things during a mobile phone-based consultation.	0.727	0.53	—	—
S408: I can understand the doctor's evaluation of my health problems.	0.757	0.57	—	—
S409: If I cannot understand the web-based doctor's explanations, I will tell them so.	0.699	0.49	—	—
Evaluating mHealth information	—	—	0.87	0.52
S601: I can evaluate the quality of health information available on my mobile phone.	0.692	0.48	—	—

Factors and items	Standard Factor Loading	Squared multiple correlations	Composite reliability	Average variance extracted
S602: I can evaluate the reliability of the evidence cited in mobile-based health information.	0.791	0.63	—	—
S603: I usually check the source of health information.	0.732	0.54	—	—
S604: I can evaluate the reliability of the source of health information.	0.772	0.6	—	—
S606: I search for health information using a variety of channels.	0.645	0.42	—	—
S702: I can identify advertisements in search results.	0.671	0.45	—	—
mHealth decision-making	—	—	0.81	0.52
S801: I am confident in applying the health information that I acquire using my mobile phone to make decisions.	0.658	0.43	—	—
S802: I believe that the decisions I make can improve my health.	0.775	0.6	—	—
S803: I incorporate my health-related decisions into my daily medical care.	0.746	0.56	—	—
S804: I can build a healthy life in accordance with the health-related decisions I make.	0.701	0.49	—	—

[a]Not applicable or data are not available.

Table 2. Discriminant validity test.

	Average variance extracted	mHealth desire	Mobile phone operational skills	Acquiring mHealth information	Acquiring mHealth services	Understanding of medical terms	Mobile-based patient–doctor communication	Evaluating mHealth information	mHealth decision-making
mHealth desire	0.538	0.733[a]	0.427	0.602	0.455	0.546	0.601	0.488	0.547
Mobile phone operational skills	0.590	0.427	0.768[a]	0.615	0.627	0.469	0.657	0.573	0.373
Acquiring mHealth information	0.655	0.602	0.615	0.809[a]	0.721	0.702	0.703	0.685	0.579
Acquiring mHealth services	0.683	0.455	0.627	0.721	0.826[a]	0.608	0.746	0.603	0.453
Understanding of medical terms	0.622	0.546	0.469	0.702	0.608	0.789[a]	0.825	0.781	0.785
Mobile-based patient–doctor communication	0.547	0.601	0.657	0.703	0.746	0.825	0.740[a]	0.762	0.733
Evaluating mHealth information	0.517	0.488	0.573	0.685	0.603	0.781	0.762	0.719[a]	0.771
mHealth decision-making	0.521	0.547	0.373	0.579	0.453	0.785	0.733	0.771	0.722[a]

[a]The square root of the average variance extracted value.

Figure 2. A comparison of the models. Statistic indices — AGFI: adjusted goodness of fit index; CFI: comparative fit index; GFI: goodness of fit index; TLI: tucker-Lewis index. Factor names — AMHI: acquiring mHealth information; AMHS: acquiring mHealth services; EMHI: evaluating mHealth information; MHD: mHealth desire; MHDM: mHealth decision-making; MPDC: mobile-based patient-doctor communication; MPOS: mobile phone operational skills; PBMHL: problem-based mHealth literacy; RMSEA: root mean square error of approximation; UMT: understanding of medical terms.

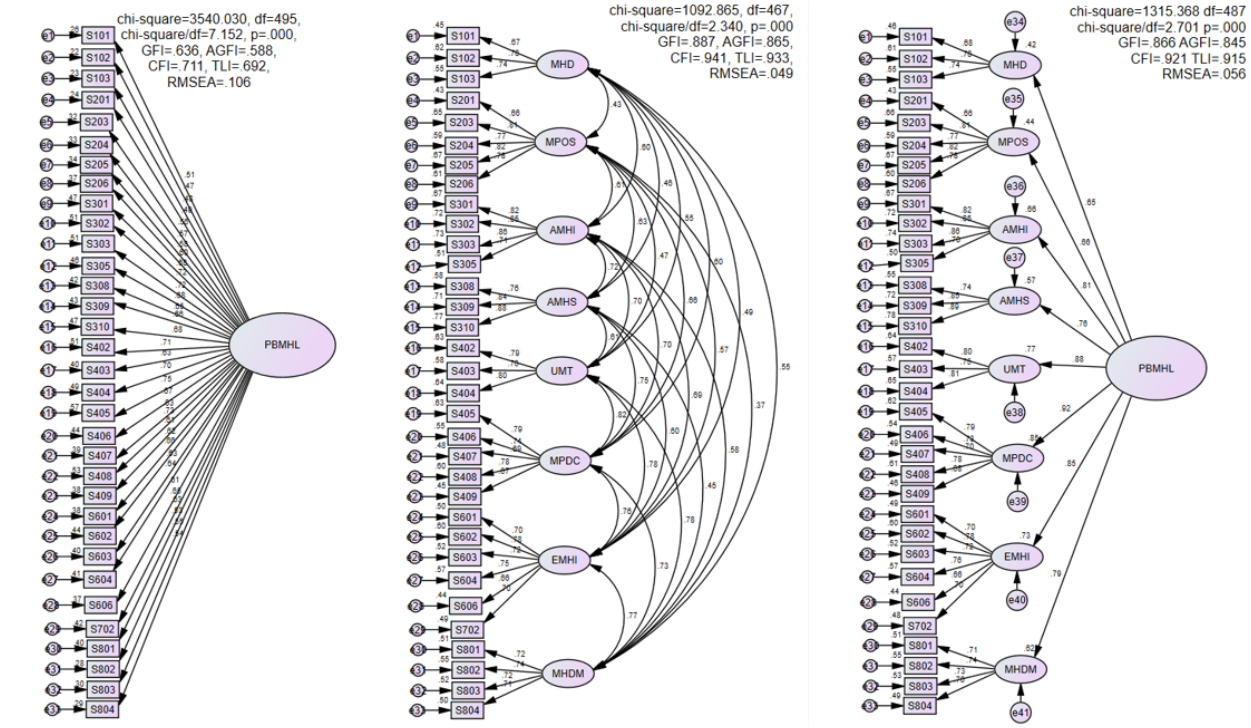


Table 3. Model comparisons.

Model comparison	Chi-square (*df*)	Chi-square to *df* ratio	GFI[a]	Adjusted GFI	Comparative fit index	Nonnormed fit index	Root mean square error of approximation
Model 1: Single-factor model	3540.030 (495)	7.152	0.636	0.588	0.711	0.692	0.106
Model 2: 8-factor correlation model	1092.865 (467)	2.34	0.887	0.865	0.941	0.933	0.049
Model 3: 8-factor second-order model	1315.368 (487)	2.701	0.866	0.845	0.921	0.915	0.056
Acceptable values	N/A[b]	<5	>0.8	>0.8	>0.9	>0.9	<0.08

[a]GFI: goodness of fit index.

[b]N/A: not applicable.

Second Validation Survey

We collected 433 valid and unique responses (men: 222/433, 51.3%; women: 211/433, 48.7%). The majority of respondents were between 30 and 60 years old (263/433, 60.8%); 50.6% (219/433) of participants possessed a middle school education or lower (senior high school: 94/433, 21.7%; college degree: 68/433, 15.7%; a bachelor's degree or above: 52/433, 12.0%), the average family monthly income was stratified as follows: 3000 RMB and below (112/433, 25.9%), 3001 to 8000 RMB (175/433, 40.4%,), 8000 to 12,000 RMB (93/433, 21.5%), and greater than 12,000 RMB (53/433, 12.2%). The proportion of participants living in urban areas was 50.3% (218/433).

The results of the exploratory factor analysis supported the feasibility of performing factor analysis (Kaiser-Meyer-Olkin 0.955; Bartlett test of sphericity: $P<.001$). As expected, the exploratory factor analysis extracted 8 factors from the 33 items; furthermore, those 8 factors cumulatively explained as much as 81.9% of the variance. The internal reliability of each factor and subscale was also within a good range, as the Cronbach α values for the 8 subscales fell between 0.864 and 0.949 and were greater than the cut-off of 0.7. Overall, these results suggested that the 8 factors extracted through the exploratory factor analysis were appropriate.

Factor loadings of all items were above 0.6, the composite reliability values were between 0.866 and0.947, and the average variance extracted values were greater than0.6. In addition, the second-order model fitness indices were in the good range (χ^2/df=2.396, GFI 0.861, adjusted GFI 0.840, comparative fit index 0.949, nonnormed fit index 0.945, RMSEA 0.057).

The result of multiple-group analysis revealed that there were no significant differences in the indices of model fit between separate second-order models for different educational levels (*P*=.31-.47) or different age groups (*P*=.29-.70) (Multimedia Appendix 3). This suggested that the PB-mHLS model was stable across different demographic groups in the sampled population.

The structural equation model to examine the impact of mHealth use on mHealth literacy and the impact of mHealth literacy on health prevention behavior met strict criteria (χ^2/df=2.251, GFI 0.860, adjusted GFI 0.840, comparative fit index 0.949, nonnormed fit index 0.945, RMSEA 0.054). The impact factor of mHealth use on mHealth literacy was 0.43, while that for mHealth literacy on health prevention behavior was 0.23. All indicators met the satisfactory significance level (*P*<.001).

Discussion

Principal Findings

Using a series of empirical studies—including an exploratory study with in-depth interviews and observations, a pilot survey using a small sample, and 2 web-based surveys using large samples—we developed and validated an mHealth literacy scale that was built on a problem-based framework. Specifically, we constructed a scale with 33 self-reported items in 8 domains and subscales to operationalize mHealth literacy. The PB-mHLS enhances our understanding of the abilities and skills required to use mobile health resources, and extends the notion of health literacy into the context of mobile communication. In this regard, mHealth literacy scales such as the PB-mHLS are likely to retain similar domains or factors that have been identified in eHealth literacy scales. The key components of eHealth literacy mainly relate to the behaviors involved in operating digital devices, and in acquiring, comprehending, as well applying (ie, making health-related decisions) digital health resources (Table 4). These components have been incorporated into the framework of the PB-mHLS. Importantly, this newly developed scale includes unique factors that reflect the distinct characteristics of mHealth. The new scale makes several important contributions to the literature on eHealth and mHealth.

First, unlike the eHEALS and the Digital Health Literacy Instrument, the PB-mHLS integrates mHealth desire as an essential aspect of eHealth practices and integrates this into the measurement of mHealth literacy. Conventional eHealth literacy scales start by evaluating individuals' behaviors of searching for health resources (eg, acquiring health information), whereas the PB-mHLS extends individuals' behavioral trajectories by assessing their desire to use the mobile internet to solve health-related problems. In other words, in the PB-mHLS, the starting point for examining individuals' health behavioral trajectories is their motivations to engage with mobile health practice. This distinction is important because a clear and strong motivation or desire for mHealth is likely to guide an individual's subsequent actions. This finding was noted in our exploratory study, where participants who were unable to appropriately understand their own health needs tended to either complicate or ignore simple health problems, resulting in worse outcomes. Indeed, when participants were at the same educational levels, those with significant health needs were found to solve their health problems more efficiently. Therefore, the ability to properly express mHealth desire should be a critical component of any mHealth literacy scale.

Table 4. Domain comparisons between the Problem-Based mHealth Literacy Scale (PB-mHLS) and eHealth literacy scales.

Domains	eHEALS (Electronic Health Literacy Scale) 1.0 and 2.0	Digital Health Literacy Instrument	Problem-based mHealth Literacy Scale
eHealth behaviors and needs	—[a]	Protecting privacy[b]	Mobile health needs
Digital device operational skills	—	Computer operational skills	Mobile phone operational skills
Acquiring health information	Information navigation awareness; information acquisition skills	Information navigation and searching	Acquiring mHealth information and services
Understanding health information	—	—	Understanding of medical terms
Online interactions	Social media interactions	Adding self-generated content	Mobile-based patient–doctor communication
Evaluation of health information	Evaluation and application of information	Evaluation of information	Evaluation of mHealth information
Application of health information	Evaluation and application of information	Application of information	mHealth decision-making

[a]Not included.

[b]This domain was not confirmed in the as its internal consistency was not acceptable (Cronbach α=0.57) and its item-total correlation was mostly nonsignificant.

Second, the PB-mHLS evaluates individuals' abilities to use digital or mobile health resources in two ways, that is, by acquiring mobile health information, and by acquiring mobile health services. To evaluate how individuals make use of digital

health resources, most eHealth literacy scales include a specific domain for the acquisition of digital health information. However, such widespread use of health information to represent health resources may be less relevant where mHealth is concerned, because individuals can access and adopt various types of real-time medical services conveniently via their mobile phones. As such, individuals' abilities to acquire health information as well as their abilities to acquire mobile health services are incorporated into the PB-mHLS. This should facilitate comprehensive assessments of individuals' capacities for using mobile health resources.

Third, the PB-mHLS emphasizes the importance of understanding how different mobile health resources are used by individuals. Just as individuals' abilities for acquiring mHealth resources include two facets, individuals' abilities for comprehending mHealth resources are separated into two aspects in the PB-mHLS, namely, the ability to understand medical terms, and the ability to engage in mobile-based patient–doctor communications. the ability to understand medical terms is considered to be an important domain in mHealth literacy because such an ability is salient for reading and understanding the acquired mHealth content, and for accessing and using common medical services such as diagnoses or medical treatments; however, mHealth services provide great opportunities for patients to consult doctors web-based in the absence of any restrictions on time and location. Hence, individual skill in patient–doctor interactions are a crucial factor facilitating mHealth practices, which is included as an important domain in the PB-mHLS.

Finally, the PB-mHLS is an instrument that is focused on assessing individuals' health-related abilities. While the literature suggests that scientific knowledge should be considered to be a component of eHealth literacy [18], the results of the confirmatory factor analysis in this study did not support such a proposition. Instead of an individual's general science literacy, an individual's ability to understand medical terms is a more specific and appropriate factor that can reflect their capacity to comprehend sophisticated medical and health-related information. Similarly, although both the eHEALS and the Digital Health Literacy Instrument included a domain that measures an individual's ability to engage in web-based or social media interactions, our findings emphasize that web-based interactions represent a focused domain; that is, the interaction between patients and doctors via mobile devices is a more relevant factor in the evaluation of mobile health literacy. In addition, the results suggested that web-based privacy is not necessarily related to mHealth literacy. This finding is consistent with those from previous research [16] and can be explained by the notion of "privacy calculus [36,37]." Although individuals are concerned about web-based and mobile privacy risks, they are willing to trade their private information for health benefits. In other words, their desire to manage and improve their health may outweigh their concerns about privacy risks. As a result, perceptions of privacy risk may influence neither health behaviors nor health literacy, and privacy web-based may not be a necessary component of either eHealth literacy or mHealth literacy.

Limitations

First, the PB-mHLS should be verified in different contexts. Although the proportion of netizens with a low-education level was accounted for in both of our survey-based studies, the proportion of netizens in rural areas remained unbalanced. Hence, to generalize the PB-mHLS, future studies should apply this scale to populations that are not skilled at using mobile phones, such as people living in rural areas, migrant workers, or low-income groups. Likewise, the development of the PB-mHLS was based on empirical data from individuals in China. Since cultural factors may influence health literacy, it is essential to validate this scale in various other cultural contexts. Second, as was the case in previous studies [16,18,19], we opted to use self-reported questionnaires in developing the PB-mHLS. This method mostly captures self-perceived or subjective responses, and thus, it cannot objectively quantify individuals' actual abilities. Therefore, future studies should design an instrument that can objectively reflect actual mHealth literacy, and investigate the similarities and differences between actual and self-reported measurements. Third, although the relationships between mHealth literacy, mHealth use, and health prevention behaviors were verified in our study, future research should use the PB-mHLS to examine the relationships between mHealth literacy and other factors, such as health outcomes, actual health practices, or overall health condition, to further validate the appropriateness of this new instrument.

Acknowledgments

This work was supported by the Ministry of Education of China's Humanities and Social Science Foundation (18YJC860049), the Philosophy and Social Science Fund of Guangdong Province (GD20XXW01) and the Major Project of The National Social Science Fund of China (21&ZD316). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the paper. In addition, we thank all the participants in this study. We would also like to thank Jiamin Zhang for her effort in the exploratory study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Details of the exploratory study.

[DOC File , 51 KB - mhealth_v10i4e31459_app1.doc]

Multimedia Appendix 2
Pilot test exploratory factor analysis.
[DOC File , 77 KB - mhealth_v10i4e31459_app2.doc]

Multimedia Appendix 3
Detailed results of the second survey study.
[DOC File , 337 KB - mhealth_v10i4e31459_app3.doc]

References

1. Statista. URL: https://www.statista.com/topics/840/smartphones/#dossierKeyfigures [accessed 2021-11-17]
2. The 48th China statistical report on internet development. China Internet Network Information Center. URL: http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/202109/P020210915523670981527.pdf [accessed 2021-11-17]
3. Free C, Phillips G, Felix L, Galli L, Patel V, Edwards P. The effectiveness of m-health technologies for improving health and health services: a systematic review protocol. BMC Res Notes 2010 Oct;3:250 [FREE Full text] [doi: 10.1186/1756-0500-3-250] [Medline: 20925916]
4. Waegemann CP. mHealth: the next generation of telemedicine? Telemed J E Health 2010 Jan;16(1):23-25. [doi: 10.1089/tmj.2010.9990] [Medline: 20070168]
5. Chib A, van Velthoven MH, Car J. mHealth adoption in low-resource environments: a review of the use of mobile healthcare in developing countries. J Health Commun 2015;20(1):4-34. [doi: 10.1080/10810730.2013.864735] [Medline: 24673171]
6. Palazuelos D, Diallo AB, Palazuelos L, Carlile N, Payne JD, Franke MF. User perceptions of an mHealth medicine dosing tool for community health workers. JMIR Mhealth Uhealth 2013 May 04;1(1):e2 [FREE Full text] [doi: 10.2196/mhealth.2459] [Medline: 25100670]
7. Chigona W, Nyemba M, Metfula A. A review on mHealth research in developing countries. J Commun Informatic 2012 Nov 28;9(2):special issue. [doi: 10.15353/joci.v9i2.3174]
8. Leventhal T, Taliaferro JP, Wong K, Hughes C, Mun S. The patient-centered medical home and health information technology. Telemed J E Health 2012 Mar;18(2):145-149. [doi: 10.1089/tmj.2011.0130] [Medline: 22304440]
9. Eysenbach G, Powell J, Kuss O, Sa ER. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002 May;287(20):2691-2700. [Medline: 12020305]
10. Ryan A, Wilson S. Internet healthcare: do self-diagnosis sites do more harm than good? Expert Opin Drug Saf 2008 May;7(3):227-229. [doi: 10.1517/14740338.7.3.227] [Medline: 18462181]
11. Schmidt K, Ernst E. Assessing websites on complementary and alternative medicine for cancer. Ann. Oncol 2004 May;15(5):733-742. [Medline: 15111340]
12. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
13. Ahmed S. mHealth literacy: characterizing people's ability to use smartphone based health-related applications. University of Illinois at Urbana-Champaign. 2017. URL: http://hdl.handle.net/2142/97267 [accessed 2022-03-31]
14. Lin T, Bautista J. Understanding the relationships between mHealth apps' characteristics, trialability, and mHealth literacy. J Health Commun 2017 Apr;22(4):346-354. [doi: 10.1080/10810730.2017.1296508] [Medline: 28323546]
15. Norman CD. eHealth literacy 2.0: problems and opportunities with an evolving concept. J Med Internet Res 2011 Dec;13(4):e125 [FREE Full text] [doi: 10.2196/jmir.2035] [Medline: 22193243]
16. van der Vaart R, Drossaert CH. Development of the digital health literacy instrument: measuring a broad spectrum of health 1.0 and health 2.0 skills. J Med Internet Res 2017 Jan 24;19(1):e27 [FREE Full text] [doi: 10.2196/jmir.6709] [Medline: 28119275]
17. Parker RM, Baker DW, Williams MV, Nurss JR. The test of functional health literacy in adults: a new instrument for measuring patients' literacy skills. J Gen Intern Med 1995 Oct;10(10):537-541. [doi: 10.1007/BF02640361] [Medline: 8576769]
18. van der Vaart R, van Deursen AJAM, Drossaert CHC, Taal E, van Dyck JAMG, van de Laar MAFJ. Does the eHealth literacy scale (eHEALS) measure what it intends to measure? validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]
19. Ma Z, Wu M. The psychometric properties of the Chinese eHealth literacy scale (C-eHEALS) in a Chinese rural population: cross-sectional validation study. J Med Internet Res 2019 Oct 22;21(10):e15720 [FREE Full text] [doi: 10.2196/15720] [Medline: 31642811]
20. Koopman RJ, Petroski GF, Canfield SM, Stuppy JA, Mehr DR. Development of the PRE-HIT instrument: patient readiness to engage in health information technology. BMC Fam Pract 2014 Jan;15:18 [FREE Full text] [doi: 10.1186/1471-2296-15-18] [Medline: 24472182]

21. Lucas RE, Baird BM. Global self-assessment. In: Eid M, Diener E, editors. Handbook of Multimethod Measurement in Psychology. Washington, DC: American Psychological Association; 2006.
22. Swann WB, Chang-Schneider C, Larsen McClarty K. Do people's self-views matter? self-concept and self-esteem in everyday life. Am Psychol 2007;62(2):84-94. [doi: 10.1037/0003-066X.62.2.84] [Medline: 17324034]
23. Ickes W, Snyder M, Garcia S. Personality influences on the choice of situations. In: Hogan R, Johnson JA, Briggs SR, editors. Handbook of Personality Psychology. San Diego, CA: Academic Press; 1997:165-195.
24. Delroy L, Paulhus SV. The self-report method. In: Robins RW, Fraley RC, Krueger RF, editors. Handbook of Research Methods in Personality Psychology. New York: Guilford; 2007:224-239.
25. Alavi C. Problem-Based Learning in a Health Sciences Curriculum. London, UK: Routledge; 1995.
26. Haith-Cooper M. Problem-based learning within health professional education. What is the role of the lecturer? A review of the literature. Nurse Educ Today 2000 May;20(4):267-272. [doi: 10.1054/nedt.1999.0397] [Medline: 10827097]
27. Churchill GA. A paradigm for developing better measures of marketing constructs. J Mark Res 1979 Feb;16(1):64-73. [doi: 10.2307/3150876]
28. Guadagnoli E, Velicer WF. Relation of sample size to the stability of component patterns. Psychol Bull 1988 Mar;103(2):265-275. [doi: 10.1037/0033-2909.103.2.265] [Medline: 3363047]
29. Carpenter S. Ten steps in scale development and reporting: a guide for researchers. Commun Methods Meas 2017 Nov 22;12(1):25-44. [doi: 10.1080/19312458.2017.1396583]
30. Goldberg L, Digman J. Revealing structure in the data: principles of exploratory factor analysis. In: Strack S, Lorr M, editors. Differentiating Normal and Abnormal Personality. New York: Springer; 1994:216-242.
31. Tinsley HEA, Tinsley DJ. Uses of factor analysis in counseling psychology research. J Couns Psychol 1987 Oct;34(4):414-424. [doi: 10.1037/0022-0167.34.4.414]
32. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Mark Res 1981 Feb;18(1):39-50. [doi: 10.2307/3151312]
33. Maruyama G. Basics of Structural Equation Modeling. Thousand Oaks, CA: Sage; 1997.
34. Doll WJ, Xia W, Torkzadeh G. A confirmatory factor analysis of the end-user computing satisfaction instrument. MIS Q 1994 Dec;18(4):453-461. [doi: 10.2307/249524]
35. Li X, Liu Q. Social media use, eHealth literacy, disease knowledge, and preventive behaviors in the COVID-19 pandemic: cross-sectional study on Chinese netizens. J Med Internet Res 2020 Oct 09;22(10):e19684 [FREE Full text] [doi: 10.2196/19684] [Medline: 33006940]
36. Youn S. Determinants of online privacy concern and its influence on privacy protection behaviors among young adolescents. J Consum Aff 2009 Sep;43(3):389-419. [doi: 10.1111/j.1745-6606.2009.01146.x]
37. Milne GR, Gordon ME. Direct mail privacy-efficiency trade-offs within an implied social contract framework. J Public Policy Mark 2018 Dec 06;12(2):206-215. [doi: 10.1177/074391569101200206]

Abbreviations

eHEALS: eHealth Literacy Scale
GFI: goodness of fit index
mHealth: mobile health
PB-mHLS: Problem-based mHealth Literacy Scale
RMB: Renminbi
RMSEA: root mean square error of approximation

Edited by L Buis; submitted 22.06.21; peer-reviewed by J Rui, W Pan; comments to author 08.10.21; revised version received 17.12.21; accepted 25.02.22; published 08.04.22.

Please cite as:
Zhang L, Li P
Problem-Based mHealth Literacy Scale (PB-mHLS): Development and Validation
JMIR Mhealth Uhealth 2022;10(4):e31459
URL: https://mhealth.jmir.org/2022/4/e31459
doi:10.2196/31459
PMID:35394446

Original Paper

A Novel Method for Evaluating Mobile Apps (App Rating Inventory): Development Study

Rachel Mackey[1*], MPH; Ann Gleason[1*], MLiS; Robert Ciulla[1*], PhD

Connected Health Branch, Defense Health Agency, Tacoma, WA, United States

*all authors contributed equally

Corresponding Author:
Ann Gleason, MLiS
Connected Health Branch
Defense Health Agency
JBLM Box 339500 MS 34
9933 West Hayes Street
Tacoma, WA, 98433-9500
United States
Phone: 1 253 278 1535
Email: ann.m.gleason3.ctr@mail.mil

Abstract

Background: Selecting and integrating health-related apps into patient care is impeded by the absence of objective guidelines for identifying high-quality apps from the many thousands now available.

Objective: This study aimed to evaluate the App Rating Inventory, which was developed by the Defense Health Agency's Connected Health branch, to support clinical decisions regarding app selection and evaluate medical and behavioral apps.

Methods: To enhance the tool's performance, eliminate item redundancy, reduce scoring system subjectivity, and ensure a broad application of App Rating Inventory–derived results, inventory development included 3 rounds of validation testing and 2 trial periods conducted over a 6-month interval. The development focused on content validity testing, dimensionality (ie, whether the tool's criteria performed as operationalized), factor and commonality analysis, and interrater reliability (reliability scores improved from 0.62 to 0.95 over the course of development).

Results: The development phase culminated in a review of 248 apps for a total of 6944 data points and a final 28-item, 3-category app rating system. The App Rating Inventory produces scores for the following three categories: evidence (6 items), content (11 items), and customizability (11 items). The final (fourth) metric is the total score, which constitutes the sum of the 3 categories. All 28 items are weighted equally; no item is considered more (or less) important than any other item. As the scoring system is binary (either the app contains the feature or it does not), the ratings' results are not dependent on a rater's nuanced assessments.

Conclusions: Using predetermined search criteria, app ratings begin with an environmental scan of the App Store and Google Play. This first step in market research funnels hundreds of apps in a given disease category down to a manageable top 10 apps that are, thereafter, rated using the App Rating Inventory. The category and final scores derived from the rating system inform the clinician about whether an app is evidence informed and easy to use. Although a rating allows a clinician to make focused decisions about app selection in a context where thousands of apps are available, clinicians must weigh the following factors before integrating apps into a treatment plan: clinical presentation, patient engagement and preferences, available resources, and technology expertise.

(JMIR Mhealth Uhealth 2022;10(4):e32643) doi:10.2196/32643

KEYWORDS

mobile health apps; app rating; app analysis methodology; app market research; mobile phone

Introduction

Background

The lack of guidelines for identifying high-quality apps from the overwhelming number of available apps creates confusion, forestalling clinical adoption. A 2019 Australian study by Byambasuren et al [1] found that two-thirds of general practitioners used mobile apps professionally but used them primarily for medical reference purposes. In the report by Byambasuren et al [1], barriers to using apps to supplement patient care included a knowledge deficiency regarding effective uses and concerns that access sources were not trustworthy. A recent review of psychological health apps noted that it falls to health care providers to evaluate the apps' literature and marketplace or follow the guidance of their colleagues or the health system [2]. The lack of guidelines and the time it takes to vet apps to find those most suited for clinical presentation have the potential to deter clinicians from integrating mobile apps into patient care and clinical practice.

Beyond a description of the app, user ratings, and testimonials, app distribution platforms neither describe an app's overall quality nor indicate whether an app can meet a clinician's needs. Descriptions posted on app stores by the software developer may be inconsistent with the app's actual content. User ratings may imply a consensus concerning an app's usability but do not necessarily reflect an app's evidence or accuracy [3]. Although a popular app may be easy to use, and usability and navigability are important considerations, it may lack therapeutic value [4].

User ratings are only moderately correlated with objective rating scales and may reflect only limited experience with an app's capabilities [5]. Pointing to the need for standardized measures, Powell et al [6] noted that someone searching for an app in the app store is likely to select the first one noticed. User ratings' ineffectiveness in qualifying an app as medically appropriate or safe has contributed to a call for a criteria-based approach that would allow for a more objective appraisal [4]. The absence of a standardized means for evaluating health apps may impede the potential for health apps to be adopted, as well as have the potential to impact patient outcomes [7]. According to Neary and Schueller [5], evaluating apps using a structured rating tool can provide users with systematic and objective information to support the informed use of such technologies.

Existing App Rating Systems

Rating guidelines are features or characteristics to consider when determining an app's viability and fit for clinical practice [5]. The authors noted the value of multidimensional ratings over those with a singular focus. Baumel et al [4] included the following components: classification (eg, intended audience), usability (eg, ease of use), visual design, user engagement (eg, an app's interactive properties), content, therapeutic persuasiveness and alliance (therapeutic rationale and relatability), a subjective evaluation of the app, credibility, and privacy and security. Password protection and the ability to import and export data, whether using the app carries potential risks, and whether technical support is available are additional considerations [6].

Oyebode et al [8] listed personalization, self-monitoring, reminders, surface credibility, social support, trustworthiness, expertise, and real-world feel as some of the most important features of health apps. Persuasive design is defined as the use of technology to change users' attitudes or behaviors and includes behavior reinforcement, behavior change, and the shaping of attitudes as success measures [9]. Of note, although having ≥1 persuasive feature is effective, incorporating too many persuasive features only increases complexity, making the app less user-friendly and possibly decreasing its effectiveness [8].

Other factors that facilitate app use include easy-to-use navigation, clear layouts and designs, and visually available health data trends, whereas barriers can be both app specific (onerous or unintuitive navigation and small font size) and user specific (lack of technology literacy, negative attitudes about technology, and lack of internet connectivity) [10]. Jeffrey et al [10] also noted the importance of educational features and customization to support user adoption and the positive impact on adoption if the app is recommended by the patient's practitioner.

A rapid review of the literature and web resources on app rating systems has revealed several standardized approaches to rating apps. Predominantly, these solutions help users select quality apps by providing a list of evaluation questions to be considered before using an app from the Google Play or Apple Store platforms. Comprehensive rating models provided by the Mobile App Rating Scale and PsyberGuide can be used to assess the usability of a mobile app [11] and assess the evidence supporting an app's content [5]. The Enlight system was developed after an extensive literature review of app rating methods. Enlight seeks to specifically evaluate the therapeutic value of health apps [4]. Other notable systems include those developed by the American Psychiatric Association (APA) [12], the Anxiety and Depression Association of America [13], and the UK National Health Service [14]. The National Health Service hosts a web-based library of approved apps, the Anxiety and Depression Association of America posts ratings of selected mental health apps on the web, and the APA has a comprehensive system for rating mental health apps that is accessible from their website. In addition, several researchers have published app rating models for specific topics. The App Quality Evaluation Tool nutrition app rating system is an example of a topical rating system [15]. The app rating models, as well as a brief description of each system, are summarized in Table 1.

Table 1. App rating systems.

Inventory or organization	Type and total items	Availability	Intended audience
ADAA[a]-reviewed mental health apps	Apps reviewed by mental health professionals based on 5 categories	Available on the ADAA website under "mobile apps"	Mental health professionals
App adviser: APA[b]	Comprehensive app evaluation model: 5 categories with 7 to 9 questions each; brief version has 8 questions total	Available on the APA [12] website	Mental health professionals
AQEL[c]	Rating scale; 51 items	Web-based questionnaire referenced in DiFilippo et al [15]	Nutrition professionals
Enlight	Research-based, comprehensive app rating system with 6 categories of rankings from very poor to very good	Tool shared in the Baumel et al [4] publication	Health professionals
HITAM[d]	Identifies factors that influence app users' acceptance of technology and behavior, such as health information seeking, social networking, and interactivity	See Kim and Park [16] for more information	App developers
MARS[e]	Professional app quality rating scale; 6 sections with 29 items; user scale has 26 items	Tool shared in the Stoyanov et al [11] publication	Health professionals
NHS[f]	App ratings conducted by experts and posted on the website	Available on the NHS digital website under "NHS Apps Library"	General audience
One Mind PsyberGuide	App ratings conducted by experts and posted on the website	Available on the One Mind PsyberGuide website	Mental health professionals

[a]ADAA: Anxiety and Depression Association of America.

[b]APA: American Psychiatric Association.

[c]AQEL: App Quality Evaluation Tool.

[d]HITAM: Health Information Technology Acceptance Model.

[e]MARS: Mobile App Rating Scale.

[f]NHS: National Health Service.

Setting the Stage for the App Rating Inventory

The Defense Health Agency's Connected Health branch is home to the research team that developed the App Rating Inventory. This branch serves as a technology resource for the Military Health System (MHS), receiving requests for mobile apps' information from providers and app developers. A standardized approach for mobile health (mHealth) market research and app evaluation is required to ensure that consistent and reliable information is provided to MHS clinicians. The research team worked with other app evaluation teams to determine whether a pre-existing tool could be modified to fit MHS needs; however, it was determined that existing tools did not meet the criteria required for use within the MHS.

To support the MHS mission, an app evaluation tool must be usable for the full spectrum of medical and behavioral conditions and be valid for use with civilian and government-developed apps. A critical requirement was that the rating system avoid subjectively defined scoring items; what was needed was an objective tool with clear and concise criteria free from personal opinion. Of equal importance was the need for a holistic accounting of each evaluated app; that is, the rating tool should include aspects that have been tested and vetted more than evidence, user experience, the value of the content; however, all 3 should be within one system. Following a review of the literature and the existing rating systems, the research team found that no existing tool met its needs.

Although disease-specific apps are the primary use case, the App Rating Inventory can also be used with nonclinical conditions (eg, activity counters, nutrition, and physical fitness). In addition to assisting clinicians from diverse disciplines with app selection, the tool is used in decision-making concerning new software development proposals and scanning the markets for similar products before committing research funds to new development.

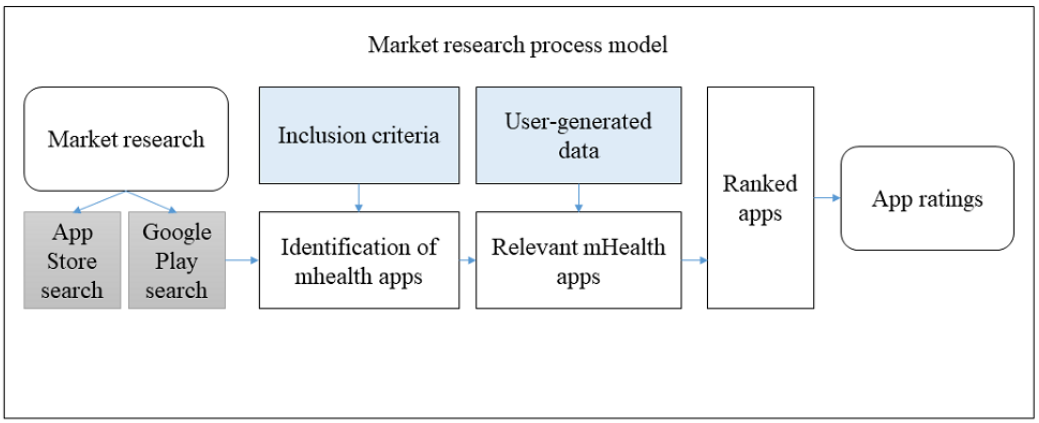
Methods

App Selection Procedure

Although a decision regarding an app's best fit for a clinical situation is supported and perhaps driven by the ratings' findings, app selections are ultimately grounded in clinical judgment. The first step in this iterative procedure is market research. In this procedure, a market search of the distribution platforms is performed before the App Rating Inventory rating system is applied. The initial market scan leads to a more detailed review of each app and its published description before alignment with the inclusion criteria, or a decision about which

apps will be rated can be determined. The protocol integrates the components of each research question. Apps that meet the inclusion criteria are funneled based on the number of criteria met. If >10 apps meet the inclusion criteria, a top-10 list is created using user-generated data from the app distribution platforms: number of user reviews, user ratings, and number of downloads, followed by the actual ratings. The process methodology for market research leading to app ratings is shown in Figure 1.

Figure 1. App rating process flowchart. mHealth: mobile health.



Early in the development process, an ANOVA was conducted to determine whether the above-described ranking process was statistically valid. The research team selected 10 top-ranked, 5 middle-ranked, and 10 bottom-ranked apps for testing with ANOVA. This was done to determine the level of variance between high- and low-ranked apps. Each of the 25 apps was rated using the App Rating Inventory, and the resulting scores were evaluated. Apps in the top-ranked grouping averaged an App Rating Inventory score of 14.77 (SD 4.63). The mean score for the middle-ranked apps was 9.66 (SD 2.25). Bottom- or lower-ranked apps received a mean App Rating Inventory score of 7.28 (SD 3.6). The ANOVA showed a statistically significant difference in the rating scores from top-ranked apps when compared with middle-ranked and bottom-ranked apps. No significant difference was observed between the middle-ranked and bottom-ranked apps. In short, the use of user-generated data to perform app rankings was found to be an effective method for selecting apps to be rated.

Development of the App Rating Inventory

Following a review of the literature and existing rating systems, it was determined that a pre-existing tool did not meet the needed requirements for use within the Defense Health Agency. An app rating tool was needed that could be used by the research team to objectively assess the quality and features of all mHealth apps, regardless of specialty. The seven subject matter experts who comprised the research team created a baseline list of the characteristics that high-quality mobile apps, which are intended for use in a clinical setting, should have. The following list was based on experience and insights from information technology staff, app developers, mHealth content experts, health care providers, and health research professionals: (1) empirical base (underlying theoretical model), (2) educational content, (3) patient-generated data, (4) interactive features, (5) entertaining and immersive, (6) user customization, (7) ease of use, and (8) free of bugs and glitches.

These eight categories served as an initial baseline upon which to build the rating system. Ongoing refinement, which was focused on operationalizing terms and eliminating ambiguity and overlap between items, produced several distinct iterations. Each new format of the inventory was piloted, tested, and subjected to an in-depth review before making additional modifications. The initial 40-item inventory was subsequently reduced to its current 28-item count.

Results

Overview

In the initial development of the App Rating Inventory, 3 rounds of testing were performed to narrow the criteria and refine the scope of the tool. The first 2 rounds of developmental tests yielded low interrater reliability (between 0.48 and 0.50). After retraining, streamlining inventory questions, and refining operational definitions, the third round of pilot testing increased the interrater reliability score to 0.62.

Following the improvement in interrater reliability, the app development team conducted the first round of analysis, implementation, and testing of the tool for 6 months, which included 96 apps rated by the research team, and the rated apps canvassed 12 distinct conditions (eg, depression, low back pain, autism, opioid use, and stress). The 2688 data points from these ratings were used for factor and commonality analyses. Validity testing was conducted following each of the 3 pilot iterations and the subsequent revisions.

During the same 6-month period, interrater reliability (2 raters) across each of the 12 topic areas was high (between 0.92 and 0.95). The inventory's now-improved internal consistencies allowed for more advanced testing. Commonality testing identified high levels of linkages among the 4 criteria, resulting in the deletion or combination of these criteria to reduce redundancy. Factor analysis resulted in the restructuring of the

linked criteria and the removal of 2 additional criteria that were identified as outliers and did not match the features in the 96 rated apps. Content validity testing illuminated weaknesses that reduced the apps' ability to perform well when administered to all mHealth apps, regardless of the topic area. The affected items were adapted to increase the consistency of all the apps. The App Rating Inventory proved to have effective utility across a broad range of clinical condition areas (eg, pain apps, substance abuse apps, and insomnia apps).

Statistical analysis and external consultation highlighted the following additional criteria of importance: privacy, peer support, emerging technology, and the encryption of exported data. The development team consulted 2 evidence-based published app evaluation owners identified by the preliminary literature review. Consultation with these expert sources was conducted both before the tools were created and during the development of the App Rating Inventory. The following criteria were adapted or added to bridge these gaps: the app connected users with social support (peer chat, social media, or support group platforms); the app included privacy settings and allowed encryption of user information and password protection; and the app used artificial intelligence (eg, chatbot and coach).

A second round of analysis tested the tool's dimensionality; that is, whether the tool's criteria were performing as operationalized. This analysis tested the predominant themes and linkages in the tool. Reliability testing was used to assess internal and external consistency. Internal analyses included rater impressions during the tool's use and tracking of the consistencies of information across research topics. Interrater reliability was evaluated throughout the testing process.

Dimensionality testing confirmed that the tool's hypothesized criteria performed as desired. Each of the tool's components reflected a unique measure. Reliability testing demonstrated that the tool performed consistently. Consistencies in ratings involving apps with disparate features and across various topics (eg, pain and insomnia) showed the tool's capacity for broad-spectrum application.

App Rating Inventory in its Final Iteration

The final App Rating Inventory was a 28-item, 3-criterion tool (see Multimedia Appendix 1 for the App Rating Inventory checklist). The scoring system changed with each iteration of the inventory, with the end goal being a more simplified procedure. In the final version, the evidence criterion contained 6 items, and the content and customizability criterion each contained 11 items. Scoring is based on a simple binary system; that is, either the app contains the feature, or it does not. The 28 items are weighted equally; no item is considered more (or less) important than any other item. Each rated app receives four scores: a score for evidence, content, and customizability and a total score (sum of the 3 categories). Higher scores indicate that the app obtained a positive score on more items than a similar app with a lower score. Evidence, content, and customizability scores allow clinicians to make focused decisions when selecting an app for clinical use. As the prevailing assumption is that clinical judgment supersedes ratings, a clinician might select an app that receives an overall lower total score than that of similar apps, as the app received a high customizability rating that especially fits the clinician's use case.

A binary approach means that raters do not have to grade their assessment along a continuum such as the systems reported in the literature that use a multipoint Likert-type scale. Using scoring for presence (rated as *1*) versus absence (rated as *0*), the App Rating Inventory minimizes subjective rater input. Although there may be value in developing broad constructs that apply to a full range of health technology platforms (*eHealth intervention programs*) [6], the App Rating Inventory's sole focus is mobile apps. A system developed to measure a mobile app's properties is not likely to equate to a website or telehealth platform.

Case Example

To search for apps that help with sleep difficulties, the distribution platforms were queried for *sleep* and *insomnia*. Extraneous results were removed from the initial findings of 1005 apps, leaving 487 (48.46%) apps. These apps were further funneled to include only apps that were free and patient focused and included sleep education; mindfulness or meditation; fatigue or risk assessment; components of cognitive behavioral therapy; and tracking to monitor sleep quantity, quality, or impact of insomnia. The final count was 8 apps that met ≥4 criteria; these apps were rated using the App Rating Inventory.

Table 2 shows the scores for 2 sleep-related apps from the overall search (in this example, both apps were developed by the federal government). There are category and total scores (the sum of the numerators for each category). Generally, only apps that achieve at least a 50% agreement threshold (scoring positively on at least 14 items) are included in the final narrative report. In the final report, detailed descriptions of the apps are accompanied by numerical ratings. The report might also include *first-hand* observations that occurred to the rater and a gap analysis when the distribution platforms did not offer apps that met the prescan inclusion criteria, especially helpful when determining whether to fund new software development.

Table 2. App Rating Inventory app rating scores.

Apps	Total App Rating Inventory score	Evidence, score out of N	Content, score out of N	Customizability, score out of N
CBT-i Coach	19	6/6	7/11	6/11
Insomnia Coach	19	5/6	7/11	7/11

Discussion

Lessons Learned

After 3 years of consistent use of the App Rating Inventory, the development team arrived at 6 fundamental observations, as discussed in the following sections.

Popularity Does Matter

Apps with a high number of downloads and user reviews (suggesting that the app is popular with users) may actually reflect app quality. Although total downloads do not ensure that an app is evidence based or has clinical utility, a high number of user reviews and associated positive ratings are signs of tangible and sustained user engagement that suggest that an app has updated, relevant quality features. In the absence of consensus resources for evaluating mHealth apps, users will choose apps with high user ratings, similar to picking one restaurant over another as it has a better star rating and later finding that it does indeed have quality food, ambience, and customer experience.

Dynamic, Interactive Content Creates Repeat Users

For apps developers, increasing app engagement is an important consideration. The repeated use of dynamic content by self-management and prevention-focused apps will increase the number of touchpoints a patient has with the associated content. These engaging features range from app reminders, pushed as notifications to the user's main device home screen, to dynamic, adaptable content that evolves as the user meets individual app goals. In the end, there is a feedback loop between app sustainment and a loyal following—loyalty incentivizes the developer to improve the app, and those improvements are rewarded by more loyalty.

The Bait and Switch Method is Common

With mobile apps, what you see is not always what you receive; in fact, there is no equivalence between distribution platforms' descriptions and what a reader can expect from a research article's abstract. The description of an app is similar to a sales pitch meant to encourage downloads. Once downloaded, the user experience may not match the marketing ploy. Perhaps, the most common occurrence is supposedly free content that the user discovers has a cost, or the user may find that the key content is locked in the free version. The user may discover that a subscription package is required; the common tagline for this is *free to download*. Indeed, the app is free to download but, once installed, cannot be used without selecting a monthly purchase plan. Although a medication management app may allow the user to enter the medication that they want to track, access to the symptom tracker, medication reminders, patient diary, calendar, and refill reminders are all separate in-app purchases or only available if the user chooses to purchase a *premium* version of the app.

There Really is an App for Everything

Most smartphone users will be familiar with the phrase "there's an app for that." From 2015 to 2017, the number of public-facing mHealth apps doubled, saturating distribution platforms with >300,000 apps [17]. The market is flooded with mHealth apps for almost every aspect of care and health. When conducting original app market research, it is common to have several hundred apps generated by each search term, regardless of the topic area. The most time-intensive step of the app rating process is the initial environmental scan, identifying alignment with inclusion criteria and determining an app's topical relevance.

Apps Can Perpetuate Inaccurate Information

App distribution platforms do not require mHealth apps to be evidence informed or supported by best practices. As apps may contain inaccurate or potentially harmful content, vetting and validating an app's clinical content before recommending it is crucial.

Mobile Apps Can Enhance the Patient Experience

The integration of mHealth apps into care has been shown to increase treatment fidelity and program adherence [18]. By increasing patient touchpoints, health literacy, and health efficacy, mHealth apps can meet the patient at where they are.

Key Considerations

The decision to recommend apps in clinical settings should be based on a comprehensive algorithm that presents diagnosis, technology literacy, app quality and content, treatment planning, accessibility and cost, and data security. Critically underlying this decision matrix is clinical judgment. Deciding which app to use may also depend on patient engagement; a low level of engagement suggests that the app should be primarily educational, whereas an app oriented toward behavior change might be more suitable for highly engaged patients [19]. Clearly, patient characteristics factor into determining an app's suitability for use in treatment or self-management [20,21].

Although a patient's input should be obtained along with the clinician's assessment of the app [3], a commentary (in the literature) that a qualified appraisal of an app's value should be grounded in the therapeutic alliance and not solely based on an objective scoring system begs the following question: are the two mutually exclusive? Powell et al [6] suggested the notion that a scoring system gets a clinician started with app selection; however, before recommending it, the responsible clinician would download and explore all of the app's features regardless of how well it scores on an objective scoring system.

The multistep app vetting process proposed by Boudreaux et al [3] includes a literature review, a search of clearinghouses and app stores, a review of app descriptions and user ratings, reviewing social media entries (both professional and patient networks), piloting the app, and obtaining feedback from patients. The case study cited in the 2014 Boudreaux et al [3] article begins with a hypothetical physician who is interested in celiac disease apps. First, the physician contacts a medical librarian to search for a systematic review. This first step in the series of aforementioned steps raises the following question: how many clinicians will have access to or sufficient free time to check in with a medical librarian? More to the point, the absence of randomized controlled trials that a literature review would reveal does not mean that useful celiac apps are not available. The physician's next steps were as follows: searching a clearinghouse, looking at user ratings in iTunes, and then

taping web-based social networks, finally leading to the physician selecting an app—at which point the physician uses the app for a day and then recommends it to a patient. These steps have value but are time intensive and, in the end, fail to obtain an independently determined, objective evaluation of the app's content. A busy clinician needs quick, digestible guidance regarding an app's merits and usefulness.

Rating systems require some initial investment in learning the scoring protocol and the system's theoretical basis. Although the App Rating Inventory research team strived to develop a system that minimized the focus on esthetic measures (potentially introducing a degree of subjectivity into the rating system), the App Rating Inventory scoring nevertheless requires initial training to best understand the meaning of the inventory's 28 items. Although the amount of time to complete a rating depends on the number and complexity of features contained within the app, experienced raters can complete an App Rating Inventory app rating in between 15 and 40 minutes.

As the App Rating Inventory was designed for use across medical conditions and with apps developed by both government and commercial vendors, the scoring system can be used outside of the MHS and by nongovernment research groups. However, it should be noted that although any clinician can use the App Rating Inventory, the inventory was developed principally for the research team's use, with the rating results reported directly to MHS providers. Although using the entire App Rating Inventory is the recommended use case, it is possible to evaluate only an app's evidence, content, or customizability depending on the clinician's needs. This type of targeted use would produce an individual score for only 1 or 2 constructs of interest. Importantly, familiarity with the App Rating Inventory can help clinicians gain insight into the components that go into a well-constructed mobile app.

However, some writers in this space argue that rating approaches that produce a score are flawed. Henson et al [20] discussed a consensus statement concerning mobile app standards, noting that a key feature of the framework (privacy and security, app effectiveness, user experience, and data integration) is that none of the framework questions are associated with a score. This concern with a system that produces an objective score is that such a score has the potential to be reductionist or somehow imply the existence of a magic formula [22]. Although there is a certain logic in the APA's stepwise hierarchical process, a clinician who has the time to categorically navigate the APA system would surely conclude that 1 app was better than similar apps. After all, the clinician must choose which among an array of apps to include in the treatment plan. It is unclear how this inevitable ranking is significantly different from the end result of an objective rating or decision matrix.

Another argument against a scoring system is that software developers are constantly making changes to apps. Excluding bug fixes, what is the evidence that developers are making constant upgrades to an app? Even bug fixes, although making an app more usable, do not necessarily alter the core content or graphics. Only a wholesale upgrade that results in an entirely new graphical interface or a navigation system renovation or removing or adding entirely new features or content would negate the results from an objective scoring system. Although content should be systematically monitored by subject matter experts involved in the app's development, what is the rate of occurrence of new medical or behavioral knowledge that would necessitate significant changes to an app's features and navigation? Consider the following for behavioral treatment apps: how often do new theoretical models emerge that would substantially alter an app intended to help with depression, stress management, or insomnia?

Perhaps, mobile apps should be subject to a certification system [6]. Although there have been attempts to certify software companies as meeting standards, there are currently no well-trusted or actively used certifications for individual apps. Although Google and Apple check for security issues, apps posted on these distribution platforms are not subjected to a certification evaluation. A centralized, curated, and easily accessed database might list both certified and uncertified apps. However, even with agreement across the industry about how a quality app is defined, it is unlikely that even the most comprehensive library will include the full spectrum of health-specific apps. Other considerations include the following: should clinicians only use apps listed in the clearinghouse; would ratings be considered obsolete after a year, necessitating a new round of evaluation; who would track updates that significantly alter the app's content; If a patient brings a noncurated app into an appointment, should the clinician discourage its use even if the patient reports that the app is beneficial; how would interrater reliability be accomplished to ensure certification accuracy; and who is the final arbiter of the app's guidelines and the associated ratings?

Selecting a best-practice app should involve no more than the following three steps: (1) query the market with key search terms; (2) check the description, user ratings, total downloads, and credibility of the developer; and (3) download and navigate the app with a particular focus on whether the content is evidence based, is founded in a theoretical model, and allows the user to input and store information (interactivity).

Should a viable clearinghouse exist, clinicians might avoid the first 2 steps; however, assuming that no clearinghouse is comprehensive, the last step is crucial. Even when the professional rating of an app is available, the last step is an essential requirement.

In summary, scoring systems provide guidance and filter down an exhaustive list of health apps in a given category to a handful for consideration. Indeed, apps are not new medicines; in many cases, they are novel delivery systems for proven interventions.

Acknowledgments

The authors would like to acknowledge the following individuals for their contributions to the App Rating Inventory: Shaunesy Walden-Behrens, MPH and MBA; Danielle Sager, MPH and MHIIM; Renee Cavanagh, PsyD; Sarah Stewart, PhD; Christina Armstrong, PhD; Julie Kinn, PhD; David Bradshaw, PhD; and Sarah Avery-Leaf, PhD.

Conflicts of Interest

None declared.

Multimedia Appendix 1
App Rating Inventory checklist.
[PDF File (Adobe PDF File), 46 KB - mhealth_v10i4e32643_app1.pdf]

References

1. Byambasuren O, Beller E, Glasziou P. Current knowledge and adoption of mobile health apps among Australian general practitioners: survey study. JMIR Mhealth Uhealth 2019 Jun 03;7(6):e13199 [FREE Full text] [doi: 10.2196/13199] [Medline: 31199343]
2. Bush NE, Armstrong CM, Hoyt TV. Smartphone apps for psychological health: a brief state of the science review. Psychol Serv 2019 May;16(2):188-195. [doi: 10.1037/ser0000286] [Medline: 30407057]
3. Boudreaux ED, Waring ME, Hayes RB, Sadasivam RS, Mullen S, Pagoto S. Evaluating and selecting mobile health apps: strategies for healthcare providers and healthcare organizations. Transl Behav Med 2014 Dec;4(4):363-371 [FREE Full text] [doi: 10.1007/s13142-014-0293-9] [Medline: 25584085]
4. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: a comprehensive quality and therapeutic potential evaluation tool for mobile and web-based eHealth interventions. J Med Internet Res 2017 Mar 21;19(3):e82 [FREE Full text] [doi: 10.2196/jmir.7270] [Medline: 28325712]
5. Neary M, Schueller SM. State of the field of mental health apps. Cogn Behav Pract 2018 Nov;25(4):531-537 [FREE Full text] [doi: 10.1016/j.cbpra.2018.01.002] [Medline: 33100810]
6. Powell AC, Torous J, Chan S, Raynor GS, Shwarts E, Shanahan M, et al. Interrater reliability of mHealth app rating measures: analysis of top depression and smoking cessation apps. JMIR Mhealth Uhealth 2016 Feb 10;4(1):e15 [FREE Full text] [doi: 10.2196/mhealth.5176] [Medline: 26863986]
7. Chan S, Torous J, Hinton L, Yellowlees P. Towards a framework for evaluating mobile mental health apps. Telemed J E Health 2015 Dec;21(12):1038-1041. [doi: 10.1089/tmj.2015.0002] [Medline: 26171663]
8. Oyebode O, Ndulue C, Alhasani M, Orji R. Persuasive mobile apps for health and wellness: a comparative systematic review. In: PERSUASIVE 2020. Cham: Springer; 2020.
9. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009 Mar;24(1):485-500. [doi: 10.17705/1CAIS.02428]
10. Jeffrey B, Bagala M, Creighton A, Leavey T, Nicholls S, Wood C, et al. Mobile phone applications and their use in the self-management of Type 2 Diabetes Mellitus: a qualitative study among app users and non-app users. Diabetol Metab Syndr 2019;11:84 [FREE Full text] [doi: 10.1186/s13098-019-0480-4] [Medline: 31636719]
11. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
12. APP advisor an American Psychiatric Association initiative. American Psychiatric Association. URL: https://www.psychiatry.org/psychiatrists/practice/mental-health-apps/why-rate-mental-health-apps [accessed 2022-03-26]
13. ADAA reviewed mental health apps. Anxiety & Depression Association of America. 2020. URL: https://adaa.org/finding-help/mobile-apps [accessed 2022-03-26]
14. NHS apps library. NHS Digital. URL: https://digital.nhs.uk/services/nhs-apps-library/guidance-for-health-app-developers-commissioners-and-assessors/how-we-assess-health-apps-and-digital-tools#step-3-technical-assessment-and-standards [accessed 2022-03-25]
15. DiFilippo KN, Huang W, Chapman-Novakofski KM. A new tool for nutrition app quality evaluation (AQEL): development, validation, and reliability testing. JMIR Mhealth Uhealth 2017 Oct 27;5(10):e163 [FREE Full text] [doi: 10.2196/mhealth.7441] [Medline: 29079554]
16. Kim J, Park HA. Development of a health information technology acceptance model using consumers' health behavior intention. J Med Internet Res 2012 Oct 01;14(5):e133 [FREE Full text] [doi: 10.2196/jmir.2143] [Medline: 23026508]
17. The growing value of digital health. The IQVIA Institute. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports/the-growing-value-of-digital-health [accessed 2022-03-27]
18. Sineath A, Lambert L, Verga C, Wagstaff M, Wingo BC. Monitoring intervention fidelity of a lifestyle behavioral intervention delivered through telehealth. Mhealth 2017;3:35 [FREE Full text] [doi: 10.21037/mhealth.2017.07.04] [Medline: 28894745]

19. Singh K, Drouin K, Newmark LP, Rozenblum R, Lee J, Landman A, et al. Developing a framework for evaluating the patient engagement, quality, and safety of mobile health applications. Issue Brief (Commonw Fund) 2016 Feb;5:1-11. [Medline: 26934758]
20. Henson P, David G, Albright K, Torous J. Deriving a practical framework for the evaluation of health apps. Lancet Digit Health 2019 Jun;1(2):e52-e54 [FREE Full text] [doi: 10.1016/S2589-7500(19)30013-5] [Medline: 33323229]
21. Torous JB, Chan SR, Gipson SY, Kim JW, Nguyen T, Luo J, et al. A hierarchical framework for evaluation and informed decision making regarding smartphone apps for clinical care. Psychiatr Serv 2018 May 01;69(5):498-500. [doi: 10.1176/appi.ps.201700423] [Medline: 29446337]
22. The Commonwealth Fund. URL: https://www.commonwealthfund.org/publications/newsletter-article/2019/jun/qa-john-torous [accessed 2022-04-06]

Abbreviations

APA: American Psychiatric Association
mHealth: mobile health
MHS: Military Health System

Edited by L Buis; submitted 04.08.21; peer-reviewed by S Pit, O Byambasuren; comments to author 14.10.21; revised version received 03.12.21; accepted 02.02.22; published 15.04.22.

Please cite as:
Mackey R, Gleason A, Ciulla R
A Novel Method for Evaluating Mobile Apps (App Rating Inventory): Development Study
JMIR Mhealth Uhealth 2022;10(4):e32643
URL: https://mhealth.jmir.org/2022/4/e32643
doi:10.2196/32643
PMID:35436227

Original Paper

Exploring Wearables to Focus on the "Sweet Spot" of Physical Activity and Sleep After Hospitalization: Secondary Analysis

S Ryan Greysen[1,2], MHS, MD; Kimberly J Waddell[2], PhD; Mitesh S Patel[3], MBA, MD

[1]Section of Hospital Medicine, University of Pennsylvania, Philadelphia, PA, United States

[2]Philadelphia Corporal Michael Crescenz Veterans Medical Center, Philadelphia, PA, United States

[3]Ascension Health, St Louis, MO, United States

Corresponding Author:
S Ryan Greysen, MHS, MD
Section of Hospital Medicine
University of Pennsylvania
3400 Spruce Street
Maloney Suite 5040
Philadelphia, PA, 19104
United States
Phone: 1 202 664 6084
Email: ryan.greysen@pennmedicine.upenn.edu

Abstract

Background: Inadequate sleep and physical activity are common during and after hospitalization, but their impact on patient-reported functional outcomes after discharge is poorly understood. Wearable devices that measure sleep and activity can provide patient-generated data to explore ideal levels of sleep and activity to promote recovery after hospital discharge.

Objective: This study aimed to examine the relationship between daily sleep and physical activity with 6 patient-reported functional outcomes (symptom burden, sleep quality, physical health, life space mobility, activities of daily living, and instrumental activities of daily living) at 13 weeks after hospital discharge.

Methods: This secondary analysis sought to examine the relationship between daily sleep, physical activity, and patient-reported outcomes at 13 weeks after hospital discharge. We utilized wearable sleep and activity trackers (Withings Activité wristwatch) to collect data on sleep and activity. We performed descriptive analysis of device-recorded sleep (minutes/night) with patient-reported sleep and device-recorded activity (steps/day) for the entire sample with full data to explore trends. Based on these trends, we performed additional analyses for a subgroup of patients who slept 7-9 hours/night on average. Differences in patient-reported functional outcomes at 13 weeks following hospital discharge were examined using a multivariate linear regression model for this subgroup.

Results: For the full sample of 120 participants, we observed a "T-shaped" distribution between device-reported physical activity (steps/day) and sleep (patient-reported quality or device-recorded minutes/night) with lowest physical activity among those who slept <7 or >9 hours/night. We also performed a subgroup analysis (n=60) of participants that averaged the recommended 7-9 hours of sleep/night over the 13-week study period. Our key finding was that participants who had both adequate sleep (7-9 hours/night) and activity (>5000 steps/day) had better functional outcomes at 13 weeks after hospital discharge. Participants with adequate sleep but less activity (<5000 steps/day) had significantly worse symptom burden (z-score 0.93, 95% CI 0.3 to 1.5; *P*=.02), community mobility (z-score –0.77, 95% CI –1.3 to –0.15; *P*=.02), and perceived physical health (z-score –0.73, 95% CI –1.3 to –0.13; *P*=.003), compared with those who were more physically active (≥5000 steps/day).

Conclusions: Participants within the "sweet spot" that balances recommended sleep (7-9 hours/night) and physical activity (>5000 steps/day) reported better functional outcomes after 13 weeks compared with participants outside the "sweet spot." Wearable sleep and activity trackers may provide opportunities to hone postdischarge monitoring and target a "sweet spot" of recommended levels for both sleep and activity needed for optimal recovery.

Trial Registration: ClinicalTrials.gov NCT03321279; https://clinicaltrials.gov/ct2/show/NCT03321279

(JMIR Mhealth Uhealth 2022;10(4):e30089) doi:10.2196/30089

KEYWORDS

sleep; physical activity; hospitalization; wearables; health care; digital health; patient reported outcomes; hospital

Introduction

Inadequate sleep and physical activity are common during and after hospitalization, but little is known about how long these disruptions continue after discharge and the consequences on functional outcomes such as mobility and overall physical health. Disruptions of normal patient activities (such as sleep, mobility, nutrition, self-care) may be traumatic for acutely ill patients and lead to a high-risk, generalized condition that has been described as "posthospital syndrome" [1-3]. The relationship between sleep and physical activity is particularly complex among patients who have been recently hospitalized [4-7]. Improving our understanding of how posthospitalization sleep and physical activity impact longer-term patient-reported outcomes will directly benefit future interventions designed to improve functional outcomes and return to normal life activities. One major challenge to studying posthospital syndrome and factors that impact sleep, physical activity, and patient-centered outcomes is the lack of data. Unlike at the hospital, the period of recovery at home after discharge is largely unobservable to providers [8-11]. Wearable sleep and activity trackers present an important opportunity to unobtrusively collect and study data during this critical and poorly understood period of care.

Previous work has examined the impact of disrupted in-hospital sleep and physical activity on short–term recovery [1-6], but much less is known about how posthospitalization sleep and physical activity impact patient-reported outcomes, especially beyond the 30-day window [12,13]. Similarly, although wearable devices have been deployed to study sleep at home, there have been very few applications of this technology to the postdischarge period specifically [14-16]. Furthermore, previous studies of the postdischarge period do not integrate patient-reported measures of sleep and activity with patient-measured wearable data to create a multidimensional, longitudinal view of this period. This lack of data on sleep and activity during the postdischarge period represents an important knowledge gap for optimizing the clinical care of hospitalized patients as well as an important opportunity for the field of mobile health (mHealth).

Given these gaps in the existing literature, we evaluated sleep and activity data from wearable devices worn by patients for 90 days (13 weeks) after hospital discharge. The purpose of this study was to examine the relationship between daily sleep and physical activity with 6 patient-reported functional outcomes at 13 weeks after hospital discharge. We used evidence-based recommendations to define ranges for healthy sleep as 7-9 hours per night [17,18] and for physical activity as over 5000 steps per day [19-21]. We explored the distribution of sleep and activity data from our study using these established parameters and then identified a subset of participants who met both guidelines for healthy sleep and activity. We defined this overlap of healthy sleep and activity as the "sweet spot" and hypothesized that this subset would have better functional outcomes compared with other participants.

Methods

Research Aims and Objectives

Our specific objectives for this study were to (1) observe the distribution of sleep and step data in our sample and describe any patterns, (2) explore differences between any observed patterns and functional outcomes (change in activities of daily living, symptom burden, physical health, quality of sleep, and life space), and (3) explore associations of functional outcomes with sleep and activity patterns using evidence-based guidelines for sleep and activity as an a priori subgroup. Our overarching hypothesis was that participants with both adequate sleep and activity would have better functional outcomes than those with only one (sleep vs activity).

Overview of Parent Trial Enrollment, Study Design, and Outcome Measures

This was a secondary analysis from the Mobility and Outcomes for Validated Evidence Incentives Trial (MOVE IT; NCT #03321279). Full details of the primary trial have been previously reported [22,23]. Briefly, this study enrolled individuals ≥18 years of age who were admitted to a general medicine or oncology service at a single, urban, academic hospital. Participants were enrolled during their hospital admission, but the intervention began after discharge. Participants were first observed for 1 week after discharge to establish a baseline level of postdischarge activity in their home. Participants were then randomized to a 12-week intervention that examined the impact of gamification and social incentives on physical activity. Gamification is the use of game design elements, such as points and levels, to motivate behavior change (such as physical activity). Social incentives, such as peer support and collaboration, can augment gamification interventions by further motivating individuals to change behaviors based on social ties [24,25]. The goal of this study was to determine if an intervention with gamification and social incentives could increase physical activity (mean steps per day) after hospital discharge. Overall, the intervention arm did not have significantly higher physical activity than the control arm, which suggests that gamification and social incentives did not alter behavior patterns after discharge for this cohort as a whole.

In addition to steps per day (which was the primary outcome of the parent trial), we assessed patient-reported outcomes at baseline (enrollment) and at 13 weeks after hospital discharge. All participants completed standardized assessments that quantified basic and instrumental activities of daily living (scales by Katz et al [26] and Lawton and Brody [27]), mobility (Life Space Assessment [28]), symptom burden (Edmonton Symptom Assessment Survey [29]), physical health (Short-Form 12 physical component scale [PCS-12] [30]), and sleep quality (Pittsburgh Sleep Quality Index [PSQI] [31]).

Ethical Considerations

The Institutional Review Board at the University of Pennsylvania approved this study (826974), and all participants

provided informed consent. Additionally, to acknowledge the time and effort contributed by participants to this study, we offered US $300 for completing all the surveys required for the study. Furthermore, we monitored closely for adverse or unexpected events and provided clear contact information to participants to reach study staff and prompt responses (within 24 hours) to any questions or concerns raised. Four participants experienced a rash on their wrist where the device was worn and were provided with cloth-based replacement bands, which resolved the issue.

Wearable Device and Research Platform Specifications

Daily steps and sleep were quantified via a wrist-worn wearable device (Withings Activité). The Activité is a wearable activity tracker designed to look and function like a wristwatch. The face of the device has hour, minute, and second hands like a traditional wristwatch with an additional dial in the center that indicates progress towards a preset step goal. The device is water-resistant to 50 meters/5 atmospheres and uses a traditional watch battery (CR2025), which does not require frequent charging and lasts about 8 months. These features help to increase consistent wear by users. Motion is detected by a high-precision MEMS 3-axis accelerometer and translated into steps or sleep using a proprietary algorithm. Synchronization between the device and user smartphone occurs automatically every 6 hours and is also triggered by each 1000 steps taken since the last sync. Previous studies by our group and others have demonstrated the reliability and validity of step counts from commercially available devices for activity studies [32], and the Activité has been validated against actigraphy and medical-grade pedometers [33].

We also leveraged the Way to Health platform, a National Institutes of Health (NIH)–funded research technology platform at the University of Pennsylvania that has been used for more than 150 clinical trials in 45 states using smartphones and wearable devices previous studies [25,34]. The Way to Health platform provides an easy way for data from participant wearable devices to be collected for research purposes through a secure server that tracks enrollment and automatically randomizes patients to intervention arms. The platform also has a messaging engine that can provide automated feedback to participants to increase the feasibility of studies (such as MOVE IT, the parent trial for this study) that use strategies such as gamification and social incentives.

Step and Sleep Data

Step data calculated by summing the number of steps in each 24-hour period (12 am to 12 am). Step data were missing in 37.7% of the current sample. Step data were considered missing for any day the participant did not wear their device, sync their data to the Way to Health platform, or reach 1000 steps. We have used this method in prior work [24,25,35,36], because evidence indicates that daily step values ≤1000 may not represent full data capture (eg, wearing the device for only a few hours in that day) [37,38]. Sensitivity analyses were conducted in primary analyses of the parent trial to explore results without step values ≤1000; no significant differences were demonstrated when step values <1000 were included. We combined multiple imputations of step data using the mice package in R (version 3.4.0), which allows for participant random effects and combined results using the standard rules by Rubin [39]. This imputation approach has been used in our prior work [24,25,36,37].

Daily sleep was calculated by summing the number of sleep minutes between 12 pm and 12 pm the following day (24-hour period). Sleep data were missing in 39.8% of the current sample. Sleep data were considered missing if the participant did not wear or sync their device, did not wear the device at night, or registered an errant number. We used multiple imputation for missing sleep data with the classification and regression trees (CART) method, which we have used in previous studies [40].

Statistical Analysis

To create our study sample for this paper from the original data set, we performed analyses in 2 steps. First, we tested for differences in the 6 patient-reported outcomes at 13 weeks by activity category (less active vs active) using all participants with available data at week 13 (n=120). Second, in an effort to disentangle the impact of sleep and physical activity on patient-reported outcomes, we then restricted the full sample to a subgroup of participants (n=60) who averaged the recommended 7-9 hours of sleep/night [17,18] over the intervention period.

The primary outcome for this analysis was the patient-reported outcomes on 6 standardized assessments administered at 13 weeks postdischarge. We calculated each participant's average daily steps and total sleep minutes over the 12-week intervention period. Participants who averaged <5000 steps/day were classified as less active, and those who averaged >5000 steps/day were classified as active [19-21]. Because all standardized assessment measures used different scoring scales, we converted all scores to z-scores.

Since our outcomes were scores from clinical assessment scales (not binary "yes/no" outcomes), we used multivariate linear regression models with each outcome as a continuous primary (dependent) variable and activity category (less active or active) as the primary predictor variable. We also adjusted for other factors as covariates, which were selected because they were different between the 2 categories at baseline: age, gender, Charlson comorbidity index, and study arm. We tested for differences in the 6 patient-reported outcomes at 13 weeks between those who were less active and those who were active but averaged between 7 hours and 9 hours of sleep per night. All analyses were completed in R [41] with a $P<.05$ significance level.

Results

Of the 232 participants in the primary trial, 120 had device-recorded sleep and step data over the entire study period and patient-reported follow-up survey data at week 13. As shown in Figure 1, patients who reported poor quality sleep (PSQI ≥5) in the month preceding hospitalization did not have different sleep patterns (mean 7.6, SD 1.8 hours per night) over the postdischarge period compared with those who had good quality sleep (mean 7.9, SD 1.7 hours per night). Similarly, we found no difference in sleep patterns for patients who were less

physically active after discharge (≤5000 steps/day; mean 7.7, SD 2.0 hours per night) compared with those who were more active (>5000 steps/day; mean 7.5, SD 1.3 hours per night; Figure 2).

As shown in Figure 3, we plotted a single mean value for sleep (minutes) and mean value for activity (steps) for each patient over the entire postdischarge period. Although we observed a linear relationship between less sleep and more activity (–2 minutes of sleep/night per +1000 steps/day; blue line in Figure 3), on closer inspection using a lens of the "T-shaped" distribution between sleep and health, we identified 4 distinct groupings: low sleep (<7 hours) and low steps (<5000), high sleep (>9 hours) and low steps (<5000), recommended sleep (7-9 hours) and low steps, and recommended sleep with high steps (>5000) or "the sweet spot."

Figure 1. Mean daily sleep by baseline (prehospitalization) Pittsburgh Sleep Quality Index (PSQI) score.

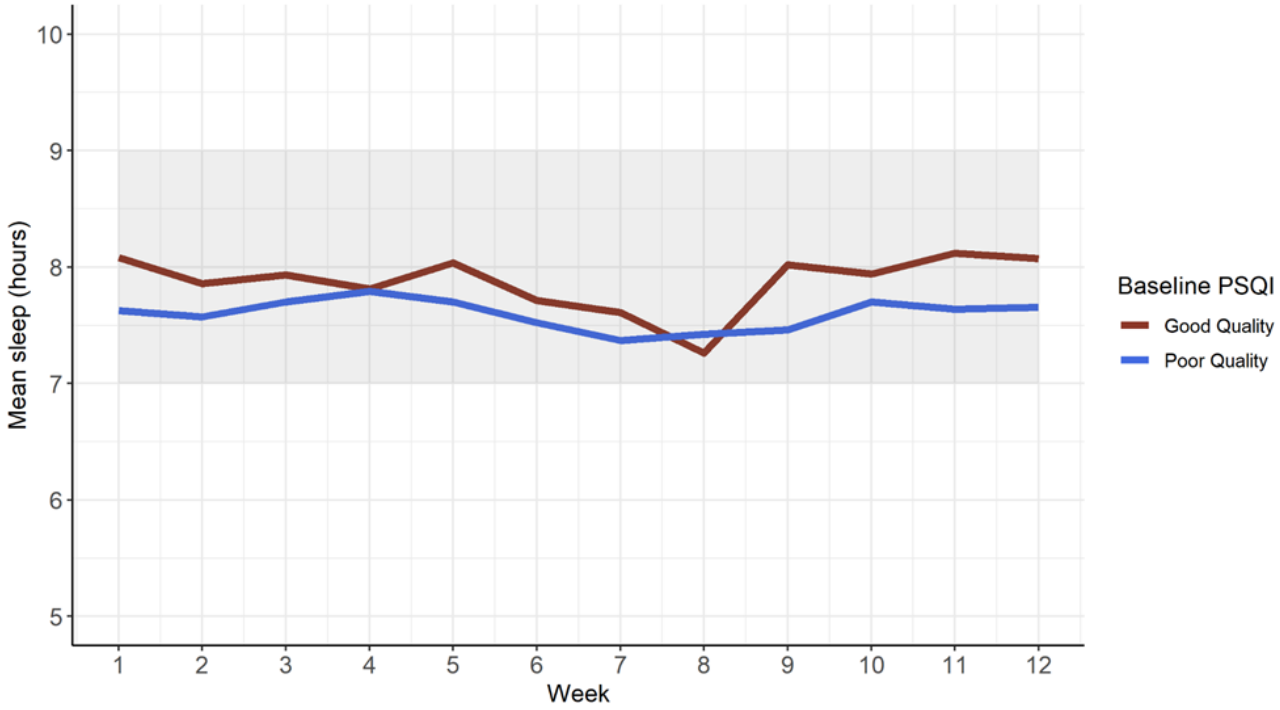


Figure 2. Mean daily sleep by level of physical activity after discharge.

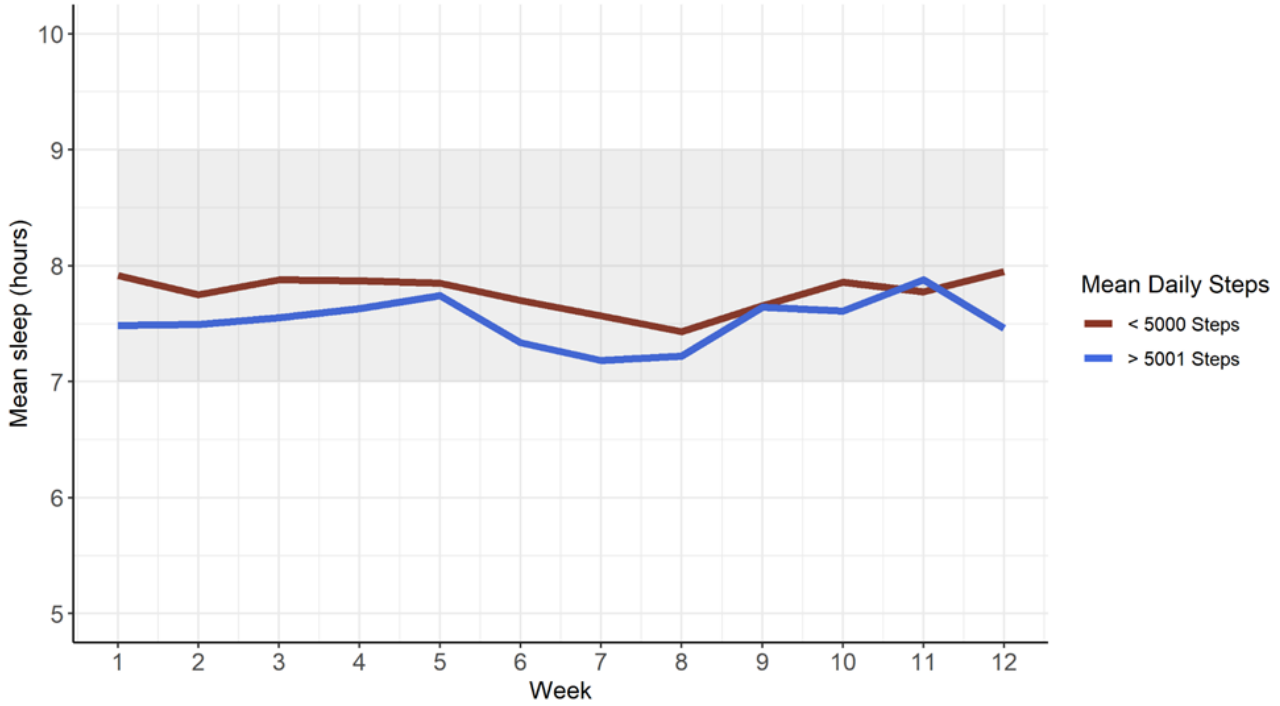
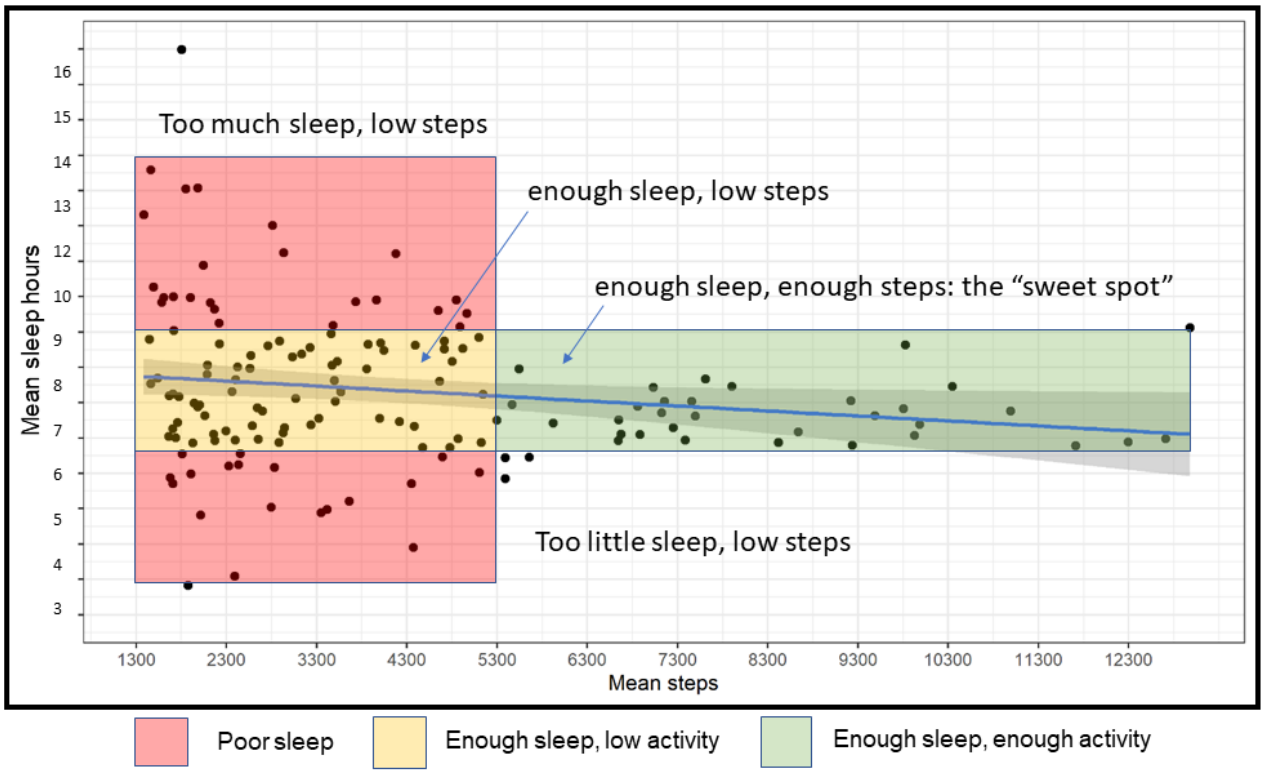


Figure 3. Mean daily sleep and mean daily activity distributions.



To further explore these clusters, we first defined participants as "less active" (<5000 steps/day; n=83) versus "active" (≥5000 steps/day; n=37). Among the 120 participants, those who were less active reported significantly worse physical health (PCS-12 z-score –0.6, 95% CI –0.9 to –0.1; *P*=.004) compared with those who were active at 13 weeks after hospital discharge. This difference, however, was present at baseline with the less active group reporting worse physical health compared with the active group (z-score mean –0.18 less active vs –0.39 active, 95% CI 0.2 to 0.9, *P*=.003). There were no other differences in patient-reported outcomes among the full sample of 120 participants.

We then focused our analysis on a subgroup of 60 participants who averaged 7 hours to 9 hours of sleep per night over 13 weeks after discharge (see clusters represented by yellow and green boxes in Figure 3). Demographics from this subgroup analysis are reported in Table 1.

Table 1. Participant demographics and baseline assessments.

Characteristics and assessments	Less active (n=36)	Active (n=24)	*P* value
Age (years), mean (SD)	41.7 (13.5)	29.8 (10.9)	<.001
Female gender, n (%)	29 (81)	10 (42)	.004
Race/ethnicity, n (%)			
Black, non-Hispanic	14 (39)	11 (46)	.92
White, non-Hispanic	17 (47)	12 (50)	
Hispanic or Latino	2 (6)	0 (0)	
Other	3 (8)	1 (4)	
Education level, n (%)			
No college degree	22 (61)	14 (58)	.99
College degree	14 (39)	10 (42)	
Annual income (US $), n (%)			
<50,000	13 (36)	9 (38)	.66
50,000-100,000	14 (39)	7 (29)	
>100,000	9 (25)	8 (33)	
Length of stay (days), mean (SD)	4.4 (2.4)	4.1 (2.1)	.56
Charlson Comorbidity Index, median (IQR)	2 [0]	1 [0]	.01
BMI (kg/m^2), mean (SD)	28.4 (8.1)	27.0 (6.8)	.48
Baseline steps per day	3032 (1201)	5417 (2561)	<.001
Baseline standardized assessments			
Activities of daily living[a]	4.7 (0.8)	4.5 (1.1)	0.61
Instrumental activities of daily living[b]	7.6 (1.1)	7.7 (1.1)	0.60
Life Space Assessment[c]	80.1 (32.1)	85.7 (27.9)	0.47
Short Form-12 physical component score[d]	34.7 (10.5)	36.0 (9.9)	0.62
Edmonton Symptom Assessment[e]	30.1 (15.2)	27.2 (13.4)	0.44
Pittsburgh Sleep Quality Index[f]	7.3 (3.8)	5.5 (3.0)	0.04

[a]Scores range from 0 to 6; higher scores indicate greater independence.

[b]Scores range from 0 to 8; higher scores indicate greater independence.

[c]Scores range from 0 to 120; higher scores indicate greater community mobility.

[d]Scores range from 0 to 100; higher scores indicate greater physical health.

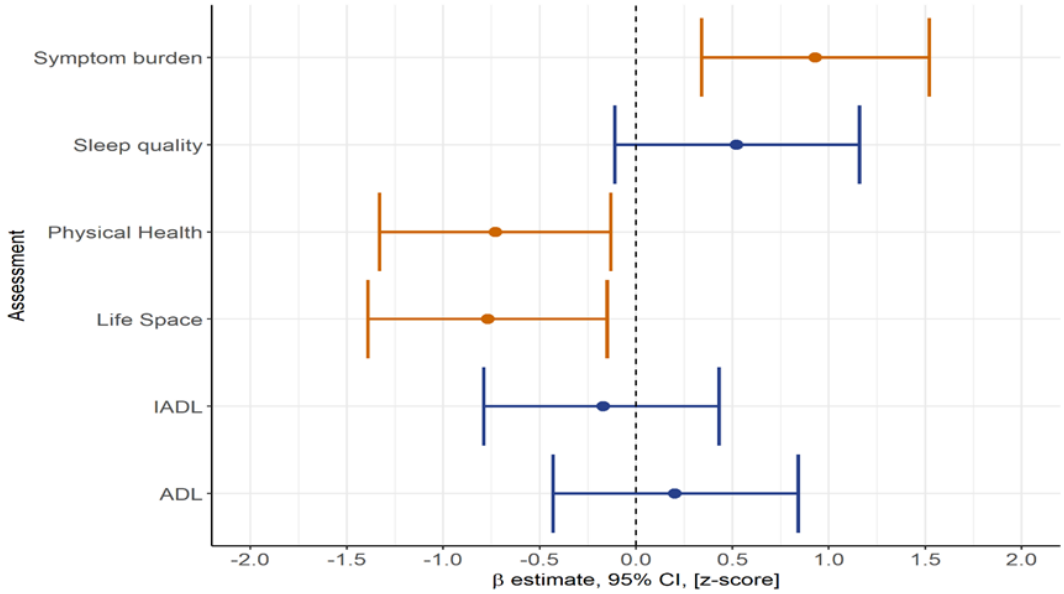
[e]Scores range from 0 to 80; higher scores indicate greater symptom burden.

[f]Scores range from 0 to 21; higher scores indicate worse sleep quality.

Less active participants (n=36) had a mean age of 41.7 (SD 13.5) years and a mean hospital length of stay of 4.4 (SD 2.4) days; in addition, 81% (29/36) were female (Table 1). Active participants (n=24) were significantly younger with a mean age of 29.8 (SD 10.9) years and a mean hospital length of stay of 4.1 (SD 2.1) days; 42% (10/24) were female. Both groups reported poor sleep quality at baseline. The less active group had a mean of 3032 (SD 1201) steps at baseline, which was significantly less than the active group, at 5417 (SD 2561) steps.

As shown in Figure 4, compared with active participants, less active participants reported significantly less life space mobility (z-score –0.77, 95% CI –1.3 to –0.15; *P*=.02), poorer physical health (z-score –0.73, 95% CI –1.3 to –0.13; *P*<.003), and significantly higher symptom burden (z-score 0.93, 95% CI 0.3 to 1.5; *P*=.02) at 13 weeks postdischarge. These differences were not observed at baseline (Table 1).

Figure 4. Beta estimates, with 95% CIs, at 13 weeks for the less active cohort; orange bars indicate significant (P<.05) comparisons, while the blue bars indicate nonsignificant comparisons. ADL: activities of daily living; IADL: instrumental activities of daily living.



Discussion

Principal Findings

Consistent with prior studies that have found a parabolic relationship [42] between sleep and worse outcomes such as cardiovascular events [43] or mortality [44], we found physical activity after hospital discharge was worst among those with either too much or too little sleep after hospital discharge. By restricting our sample to exclude those with too much or too little and focusing on those who slept the recommended amount (average 7-9 hours/night) [32,33], we were able to disentangle sleep and physical activity as an initial step to better understand the "sweet spot" of these 2 behaviors on longer-term patient-reported outcomes. In a subgroup of participants who averaged the recommended 7-9 hours of sleep a night, those who were less active (<5000 steps/day) reported significantly reduced life space mobility, physical health, and increased symptom burden, compared with those who were active (>5000 steps/day), at 13 weeks after hospitalization. These significant differences were not observed at baseline but emerged over time, suggesting that prolonged, low levels of physical activity over 13 weeks after discharge from the hospital are associated with worse patient-reported outcomes and may hinder functional recovery. Conventional wisdom might suggest that patients recovering from hospitalization should focus on rest to maximize functional recovery (ie, sleep 8-10 hours rather than 7-9 hours and walk less than 5000 steps), but our results suggest a shift in focus: It may be more important to focus on the right balance of sleep and activity rather than to maximize rest.

Our findings are especially striking considering that our sample was relatively young (mean age 40 years) and healthy (Charlson Comorbidity Index score range: 1-2) compared with a typical population of adults hospitalized for general medicine services. Although this likely represents selection bias (see the Limitations section) for our sample, we were surprised by these findings in a relatively young and healthy cohort. The relationship between sedentary levels of physical activity and adverse outcomes in older, multimorbid adults is very well-known, especially during and after acute illness [45,46]. Indeed, posthospital syndrome was first described in the Medicare population, and low mobility in this population has been described as "toxic" [47] and "epidemic" [48]. It is thus likely that the effects described here may occur at even lower thresholds, below 5000 steps/day, in older adults with greater comorbidity. This represents an important and testable hypothesis for future studies leveraging mHealth technologies. Furthermore, although studies with older adults have not suggested any differential impact of low levels of physical activity by gender, we observed that patients who were less active in the postdischarge period were more likely to be middle-aged women. We did not observe differences by race/ethnicity, income, education, or BMI. Our findings suggest that posthospital syndrome may impact younger and healthier patients who may be considered low risk by clinicians, and middle-aged women in particular may be at higher risk of being overlooked. These are also testable hypotheses for future mHealth studies with larger enrollment.

These findings add to the growing body of literature exploring the effects of sleep and physical activity on functional outcomes after hospital discharge. Previous work has primarily focused on older adults or critically ill patients [4,5,13,49], while this study's population was younger and admitted to a general medicine or oncology service. Compared with older adults and critically ill patients, those who are younger and without critical illness may have greater ability to return to normal sleep and activity patterns after discharge. However, in our cohort of younger adults without critical illness and adequate sleep, sedentary activity appears to be associated with worse functional outcomes. This provides empirical data to support the application of geriatric paradigms for sleep, activity, and other

aspects of transitions of care to a broader (nongeriatric) population [50,51].

Limitations

Several limitations should be considered when interpreting these data. We examined patient-reported outcomes at a single time point. First, given that our sample was younger and healthier than the general population of hospitalized patients, it is likely that older and sicker patients opted to wear their devices less consistently, thus limiting our ability to describe the effects of poor sleep or low activity levels on this vulnerable population. Discontinuation of device use among high-risk populations after discharge has been observed in other studies and represents an important challenge for the field of mHealth [52]. Second, these cross-sectional findings do not provide insight into when the observed differences in patient-reported outcomes emerged over the 13-week period, and future work may want to investigate this understudied temporal relationship. Third, the small sample size limits generalizability; therefore, our results should be viewed as hypothesis-generating and inform future, larger studies examining these complex constructs. Fourth, we did not have in-hospital sleep and activity data for this sample, limiting our ability to examine the impact of in-hospital sleep and physical activity on longer-term outcomes and the potential contribution to recovery. Last, device-measured prehospital activity levels (steps per day) and sleep levels (hours per night) were unknown, and without these data, we were unable to examine if participants' posthospitalization sleep and activity patterns differed from prehospitalization patterns.

Conclusion

In conclusion, these findings are an important step in understanding how posthospitalization sleep and physical activity impact longer-term functional outcomes. Participants within the "sweet spot" that balances recommended sleep (7-9 hours/night) and physical activity (>5000 steps/day) reported better functional outcomes after 13 weeks compared with participants outside the "sweet spot." Future interventions to improve functional outcomes posthospitalization should leverage wearable devices to further explore the effects of the "sweet spot" of sleep and activity on functional outcomes.

Conflicts of Interest

MSP is the owner of Catalyst Health LLC, a behavior change and technology consulting company. MSP is also an advisory board member for Humana and GlaxoSmithKline .

References

1. Krumholz HM. Post-hospital syndrome — an acquired, transient condition of generalized risk. N Engl J Med 2013 Jan 10;368(2):100-102. [doi: 10.1056/nejmp1212324]
2. Detsky AS, Krumholz HM. Reducing the trauma of hospitalization. JAMA 2014 Jun 04;311(21):2169-2170. [doi: 10.1001/jama.2014.3695] [Medline: 24788549]
3. Goldwater DS, Dharmarajan K, McEwan BS, Krumholz HM. Is posthospital syndrome a result of hospitalization-induced allostatic overload? J Hosp Med 2018 May 30;13(5):1. [doi: 10.12788/jhm.2986] [Medline: 29813141]
4. Stewart NH, Arora VM. Sleep in hospitalized older adults. Sleep Med Clin 2018 Mar;13(1):127-135 [FREE Full text] [doi: 10.1016/j.jsmc.2017.09.012] [Medline: 29412979]
5. Beveridge C, Knutson K, Spampinato L, Flores A, Meltzer DO, Van Cauter E, et al. Daytime physical activity and sleep in hospitalized older adults: association with demographic characteristics and disease severity. J Am Geriatr Soc 2015 Jul;63(7):1391-1400 [FREE Full text] [doi: 10.1111/jgs.13520] [Medline: 26131982]
6. Morse A, Bender E. Sleep in hospitalized patients. Clocks Sleep 2019 Mar 25;1(1):151-165 [FREE Full text] [doi: 10.3390/clockssleep1010014] [Medline: 33089161]
7. Rawal S, Kwan JL, Razak F, Detsky AS, Guo Y, Lapointe-Shaw L, et al. Association of the trauma of hospitalization with 30-day readmission or emergency department visit. JAMA Intern Med 2019 Jan 01;179(1):38-45 [FREE Full text] [doi: 10.1001/jamainternmed.2018.5100] [Medline: 30508018]
8. Wachter RM, Auerbach AD. Filling the black hole of hospital discharge (editorial in response to article by Walz et al.,J Gen Intern Med 2011). J Gen Intern Med 2011 Apr;26(4):354-355 [FREE Full text] [doi: 10.1007/s11606-011-1643-7] [Medline: 21327528]
9. Greysen SR, Schiliro D, Horwitz LI, Curry L, Bradley EH. "Out of sight, out of mind": housestaff perceptions of quality-limiting factors in discharge care at teaching hospitals. J Hosp Med 2012;7(5):376-381 [FREE Full text] [doi: 10.1002/jhm.1928] [Medline: 22378723]
10. Greysen SR, Hoi-Cheung D, Garcia V, Kessell E, Sarkar U, Goldman L, et al. "Missing pieces"--functional, social, and environmental barriers to recovery for vulnerable older adults transitioning from hospital to home. J Am Geriatr Soc 2014 Aug;62(8):1556-1561 [FREE Full text] [doi: 10.1111/jgs.12928] [Medline: 24934494]
11. Cain CH, Neuwirth E, Bellows J, Zuber C, Green J. Patient experiences of transitioning from hospital to home: an ethnographic quality improvement project. J Hosp Med 2012;7(5):382-387. [doi: 10.1002/jhm.1918] [Medline: 22378714]
12. Kessler R, Knutson KL, Mokhlesi B, Anderson SL, Shah M, Meltzer DO, et al. Sleep and activity patterns in older patients discharged from the hospital. Sleep 2019 Oct 21;42(11):1 [FREE Full text] [doi: 10.1093/sleep/zsz153] [Medline: 31310317]

13. Altman MT, Knauert MP, Pisani MA. Sleep disturbance after hospitalization and critical illness: a systematic review. Ann Am Thorac Soc 2017 Sep;14(9):1457-1468 [FREE Full text] [doi: 10.1513/AnnalsATS.201702-148SR] [Medline: 28644698]
14. Guillodo E, Lemey C, Simonnet M, Walter M, Baca-García E, Masetti V, HUGOPSY Network, et al. Clinical applications of mobile health wearable-based sleep monitoring: systematic review. JMIR Mhealth Uhealth 2020 Apr 01;8(4):e10733 [FREE Full text] [doi: 10.2196/10733] [Medline: 32234707]
15. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Accuracy of wristband Fitbit models in assessing sleep: systematic review and meta-analysis. J Med Internet Res 2019 Nov 28;21(11):e16273 [FREE Full text] [doi: 10.2196/16273] [Medline: 31778122]
16. Degroote L, Hamerlinck G, Poels K, Maher C, Crombez G, De Bourdeaudhuij I, et al. Low-cost consumer-based trackers to measure physical activity and sleep duration among adults in free-living conditions: validation study. JMIR Mhealth Uhealth 2020 May 19;8(5):e16674 [FREE Full text] [doi: 10.2196/16674] [Medline: 32282332]
17. Hirshkowitz M, Whiton K, Albert SM, Alessi C, Bruni O, DonCarlos L, et al. National Sleep Foundation's updated sleep duration recommendations: final report. Sleep Health 2015 Dec;1(4):233-243. [doi: 10.1016/j.sleh.2015.10.004] [Medline: 29073398]
18. Watson NF, Badr MS, Belenky G, Bliwise DL, Buxton OM, Buysse D, et al. Recommended Amount of Sleep for a Healthy Adult: A Joint Consensus Statement of the American Academy of Sleep Medicine and Sleep Research Society. Sleep 2015 Jun 01;38(6):843-844 [FREE Full text] [doi: 10.5665/sleep.4716] [Medline: 26039963]
19. Tudor-Locke C, Craig CL, Thyfault JP, Spence JC. A step-defined sedentary lifestyle index: <5000 steps/day. Appl Physiol Nutr Metab 2013 Feb;38(2):100-114 [FREE Full text] [doi: 10.1139/apnm-2012-0235] [Medline: 23438219]
20. Physical Activity Guidelines for Americans. US Department of Health and Human Services. URL: https://health.gov/paguidelines [accessed 2022-03-28]
21. Physical Activity Guidelines for Americans, 2nd edition. U.S. Department of Health and Human Services. URL: https://health.gov/sites/default/files/2019-09/Physical_Activity_Guidelines_2nd_edition.pdf [accessed 2022-03-28]
22. Greysen HM, Reale C, Mercede A, Patel MS, Small D, Snider C, et al. Mobility and outcomes for validated evidence - Incentive trial (MOVE IT): Randomized clinical trial study protocol. Contemp Clin Trials 2020 Feb;89:105911. [doi: 10.1016/j.cct.2019.105911] [Medline: 31838257]
23. Greysen SR, Changolkar S, Small DS, Reale C, Rareshide CAL, Mercede A, et al. Effect of behaviorally designed gamification with a social support partner to increase mobility after hospital discharge: a randomized clinical trial. JAMA Netw Open 2021 Mar 01;4(3):e210952 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.0952] [Medline: 33760089]
24. Patel MS, Small DS, Harrison JD, Fortunato MP, Oon AL, Rareshide CAL, et al. Effectiveness of behaviorally designed gamification interventions with social incentives for increasing physical activity among overweight and obese adults across the United States: the STEP UP randomized clinical trial. JAMA Intern Med 2019 Dec 01;179(12):1624-1632 [FREE Full text] [doi: 10.1001/jamainternmed.2019.3505] [Medline: 31498375]
25. Patel MS, Benjamin EJ, Volpp KG, Fox CS, Small DS, Massaro JM, et al. Effect of a game-based intervention designed to enhance social incentives to increase physical activity among families: the BE FIT randomized clinical trial. JAMA Intern Med 2017 Nov 01;177(11):1586-1593 [FREE Full text] [doi: 10.1001/jamainternmed.2017.3458] [Medline: 28973115]
26. Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW. Studies of illness in the aged. The index of ADL a standardized measure of biological and psychosocial function. JAMA 1963 Sep 21;185:914-919. [doi: 10.1001/jama.1963.03060120024016] [Medline: 14044222]
27. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. The Gerontologist 1969 Sep 01;9(3 Part 1):179-186. [doi: 10.1093/geront/9.3_part_1.179]
28. Peel C, Sawyer Baker P, Roth DL, Brown CJ, Brodner EV, Allman RM. Assessing mobility in older adults: the UAB Study of Aging Life-Space Assessment. Phys Ther 2005 Oct;85(10):1008-1119. [Medline: 16180950]
29. Bruera E, Kuehn N, Miller MJ, Selmser P, Macmillan K. The Edmonton Symptom Assessment System (ESAS): a simple method for the assessment of palliative care patients. J Palliat Care 1991;7(2):6-9. [Medline: 1714502]
30. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [doi: 10.1097/00005650-199603000-00003] [Medline: 8628042]
31. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
32. Case MA, Burwick HA, Volpp KG, Patel MS. Accuracy of smartphone applications and wearable devices for tracking physical activity data. JAMA 2015 Feb 10;313(6):625-626. [doi: 10.1001/jama.2014.17841] [Medline: 25668268]
33. Au-Yeung WM, Kaye JA, Beattie Z. Step count standardization: validation of step counts from the Withings Activite using PiezoRxD and wGT3X-BT. Annu Int Conf IEEE Eng Med Biol Soc 2020 Jul;2020:4608-4611 [FREE Full text] [doi: 10.1109/EMBC44109.2020.9176511] [Medline: 33019020]
34. Way to Health. URL: https://www.waytohealth.org/ [accessed 2022-03-28]
35. Chokshi NP, Adusumalli S, Small DS, Morris A, Feingold J, Ha YP, et al. Loss-framed financial incentives and personalized goal-setting to increase physical activity among ischemic heart disease patients using wearable devices: the ACTIVE

REWARD randomized trial. J Am Heart Assoc 2018 Jun 13;7(12):1 [FREE Full text] [doi: 10.1161/JAHA.118.009173] [Medline: 29899015]
36. Patel MS, Asch DA, Rosin R, Small DS, Bellamy SL, Heuer J, et al. Framing financial incentives to increase physical activity among overweight and obese adults: a randomized, controlled trial. Ann Intern Med 2016 Mar 15;164(6):385-394 [FREE Full text] [doi: 10.7326/M15-1635] [Medline: 26881417]
37. Bassett DR, Wyatt HR, Thompson H, Peters JC, Hill JO. Pedometer-measured physical activity and health behaviors in U.S. adults. Med Sci Sports Exerc 2010 Oct;42(10):1819-1825 [FREE Full text] [doi: 10.1249/MSS.0b013e3181dc2e54] [Medline: 20305579]
38. Kang M, Rowe DA, Barreira TV, Robinson TS, Mahar MT. Individual information-centered approach for handling physical activity missing data. Res Q Exerc Sport 2009 Jun;80(2):131-137. [doi: 10.1080/02701367.2009.10599546] [Medline: 19650377]
39. Rubin DB. Multiple imputation for nonresponse in surveys. New York, NY: John Wiley & Sons, Ltd; 1987. URL: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470316696.fmatter [accessed 2022-03-28]
40. Waddell KJ, Changolkar S, Szwartz G, Godby S, Patel MS. Association of behavioral phenotypes with changes in sleep during a workplace wellness program. Am J Health Promot 2021 Nov;35(8):1061-1070. [doi: 10.1177/08901171211015089] [Medline: 33998296]
41. A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. 2019. URL: https://www.R-project.org [accessed 2022-03-28]
42. Bliwise DL, Young TB. The parable of parabola: what the U-shaped curve can and cannot tell us about sleep. Sleep 2007 Dec;30(12):1614-1615 [FREE Full text] [doi: 10.1093/sleep/30.12.1614] [Medline: 18246971]
43. Wang Y, Wang J, Chen S, Li J, Lu Q, Vitiello MV, et al. Association of longitudinal patterns of habitual sleep duration with risk of cardiovascular events and all-cause mortality. JAMA Netw Open 2020 May 01;3(5):e205246 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.5246] [Medline: 32442289]
44. Liu T, Xu C, Rota M, Cai H, Zhang C, Shi M, et al. Sleep duration and risk of all-cause mortality: A flexible, non-linear, meta-regression of 40 prospective cohort studies. Sleep Med Rev 2017 Apr;32:28-36. [doi: 10.1016/j.smrv.2016.02.005] [Medline: 27067616]
45. Song J, Lindquist LA, Chang RW, Semanik PA, Ehrlich-Jones LS, Lee J, et al. Sedentary behavior as a risk factor for physical frailty independent of moderate activity: results from the Osteoarthritis Initiative. Am J Public Health 2015 Jul;105(7):1439-1445. [doi: 10.2105/AJPH.2014.302540] [Medline: 25973826]
46. Semanik PA, Lee J, Song J, Chang RW, Sohn M, Ehrlich-Jones LS, et al. Accelerometer-monitored sedentary behavior and observed physical function loss. Am J Public Health 2015 Mar;105(3):560-566. [doi: 10.2105/AJPH.2014.302270] [Medline: 25602883]
47. Greysen SR, Patel MS. Annals for Hospitalists Inpatient Notes - bedrest is toxic-why mobility matters in the hospital. Ann Intern Med 2018 Jul 17;169(2):HO2-HO3. [doi: 10.7326/M18-1427] [Medline: 30014112]
48. Brown CJ, Redden DT, Flood KL, Allman RM. The underrecognized epidemic of low mobility during hospitalization of older adults. J Am Geriatr Soc 2009 Sep;57(9):1660-1665. [doi: 10.1111/j.1532-5415.2009.02393.x] [Medline: 19682121]
49. Solverson KJ, Easton PA, Doig CJ. Assessment of sleep quality post-hospital discharge in survivors of critical illness. Respir Med 2016 May;114:97-102 [FREE Full text] [doi: 10.1016/j.rmed.2016.03.009] [Medline: 27109818]
50. Arbaje AI, Kansagara DL, Salanitro AH, Englander HL, Kripalani S, Jencks SF, et al. Regardless of age: Incorporating principles from geriatric medicine to improve care transitions for patients with complex needs. J Gen Intern Med 2014 Jun;29(6):932-939 [FREE Full text] [doi: 10.1007/s11606-013-2729-1] [Medline: 24557511]
51. Arbaje AI, Wolff JL, Yu Q, Powe NR, Anderson GF, Boult C. Postdischarge environmental and socioeconomic factors and the likelihood of early hospital readmission among community-dwelling Medicare beneficiaries. Gerontologist 2008 Aug;48(4):495-504. [doi: 10.1093/geront/48.4.495] [Medline: 18728299]
52. Patel MS, Polsky D, Kennedy EH, Small DS, Evans CN, Rareshide CAL, et al. Smartphones vs wearable devices for remotely monitoring physical activity after hospital discharge: a secondary analysis of a randomized clinical trial. JAMA Netw Open 2020 Feb 05;3(2):e1920677 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.20677] [Medline: 32031643]

Abbreviations

CART: classification and regression tree
mHealth: mobile health
MOVE IT: Mobility and Outcomes for Validated Evidence Incentives Trial
PCS-12: Short-Form 12 physical component scale
PSQI: Pittsburgh Sleep Quality Index

Edited by L Buis; submitted 30.07.21; peer-reviewed by N Koenders, S Bhattacharjee; comments to author 07.10.21; revised version received 15.12.21; accepted 18.02.22; published 27.04.22.

Please cite as:
Greysen SR, Waddell KJ, Patel MS
Exploring Wearables to Focus on the "Sweet Spot" of Physical Activity and Sleep After Hospitalization: Secondary Analysis
JMIR Mhealth Uhealth 2022;10(4):e30089
URL: https://mhealth.jmir.org/2022/4/e30089
doi:10.2196/30089
PMID:35476034

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org