JMIR mHealth and uHealth

Impact Factor (2024): 5.4
Volume 9 (2021), Issue 3 ISSN 2291-5222 Editor in Chief: Lorraine Buis, PhD, MSI

Contents

Tutorials

Corrigenda and Addenda

Viewpoint

Digital Health Integration Assessment and Maturity of the United States Biopharmaceutical Industry: Forces Driving the Next Generation of Connected Autoinjectable Devices

Ramin Rafiei[1*], PhD; Chelsea Williams[1*], MPH; Jeannette Jiang[1*], BSc; Timothy Dy Aungst[2], PharmD; Matthias Durrer[1*], BSc; Dao Tran[1*], PharmD; Ralph Howald[1*], MSc

[1]SHL Medical, Zug, Switzerland

[2]Department of Pharmacy Practice, MCPHS University, Worcester, MA, United States

[*]these authors contributed equally

Corresponding Author:
Chelsea Williams, MPH
SHL Medical
Gubelstrasse 22, 6300
Zug
Switzerland
Phone: 1 5617135654
Email: chelsea.williams@shl-medical.com

Abstract

Autoinjectable devices continue to provide real-life benefits for patients with chronic conditions since their widespread adoption 30 years ago with the rise of macromolecules. Nonetheless, issues surrounding adherence, patient administration techniques, disease self-management, and data outcomes at scale persist despite product design innovation. The interface of drug device combination products and digital health technologies formulates a value proposition for next-generation autoinjectable devices to power the delivery of precision care at home and achieve the full potential of biologics. Success will largely be dependent on biopharma's digital health maturity to implement this framework. This viewpoint measures the digital health maturity of the top 15 biopharmaceutical companies in the US biologics autoinjector market and establishes the framework for next-generation autoinjectable devices powering home-based precision care and the need for formal digital health training.

(JMIR Mhealth Uhealth 2021;9(3):e25406) doi:10.2196/25406

KEYWORDS

digital health; artificial intelligence; drug delivery; biopharma; autoinjector; injectable devices; disease management; autoimmune; oncology; rare diseases

Introduction

The interface of drug device combination products (DDCPs) and digital health technologies is rapidly expanding to provide new and innovative ways to improve a patient's health care outcomes. DDCPs are therapeutic and diagnostic products that combine drugs, devices, or biological products and include prefilled syringes or autoinjectors. The digital health frontier is manifesting in many forms, such as software as a medical device, regulated wearable devices, and telemedicine and remote patient monitoring [1]. Although many definitions of digital health have been published, overall, these definitions encapsulate empowering patients and providers with technology that can lead to scalable medical care leveraging novel digital tools [2-4]. Progress toward embracing digital health has been sporadic over the past decade, and biopharmaceutical companies are no exception [5]. As they adopt digital health, they will also need to account for the inevitable shift of the US health care environment as it gravitates toward treatment in the home, focus on patient preferences, expanded outcomes, biosimilar adoption, and broader value-based care agreements [6,7]. Sophisticated digital health technologies can measure and monitor patient outcomes, address gaps in patient care, and support medication optimization; however, demonstration of their value will require the generation of clinical, economic, and usability evidence using data resources, predictive analytics, expanded endpoints (eg, digital biomarkers), and behavioral sciences, often superseding conventional models [8].

Digital health is already demonstrating potential when combined with DDCPs [9]. Autoinjectable devices have provided real-life

benefits for patients in terms of drug self-administration, since their widespread adoption began 30 years ago with the rise of macromolecules for chronic conditions in the autoimmune space [10]. The share of biological products, many of which could be administered by autoinjectors, is growing in the US market, and they accounted for more than one-fourth of all new molecular entities approved (2015-2019) [11]. The evolution of how these drugs are delivered to patients has enabled companies to capture and drive market share and to create high brand loyalty. In recent years, there have been considerable shifts of design to enable patients to more easily utilize the device in their home setting by reducing the number of steps for activation and self-administration [12,13]. Nonetheless, issues surrounding adherence, patient administration techniques, disease self-management, and data outcomes at scale still persist despite product design innovation, and these are the next areas to be explored [14-18]. Arguably, this is a systems-based issue that affects multiple stakeholders beyond biopharma and is yet to be adequately addressed and solved.

Patients with chronic conditions are estimated to be nonadherent to their medications 25%-50% of the time, and those requiring home injections are no exception [19-25]. Majority of these biologics are administered via autoinjectors, which help improve patient adherence and are preferred for subcutaneous self-administration [26]. Research has identified that patients who are nonadherent to their autoinjectors incur high health care spending and exhibit further disease progression [22,26]. Digital health has been explored as a possible solution to this problem [27]. Biopharma has been an advocate for integrating digital technologies to address nonadherence, but there has been a considerable lack of advancement when looking at the injectable space. This slow pace of digital health adoption has often been attributed to regulatory barriers [28], although federal regulators are redefining their models for the evaluation of digital technologies, facilitating adoption [29].

Considering the expanse of the biologics market, the popularity of autoinjectors for patient self-administration, and the potential of digital health technologies advancing the role of autoinjectors in chronic disease management, we evaluated biopharma's digital health maturity as an enabler of next-generation autoinjectable devices.

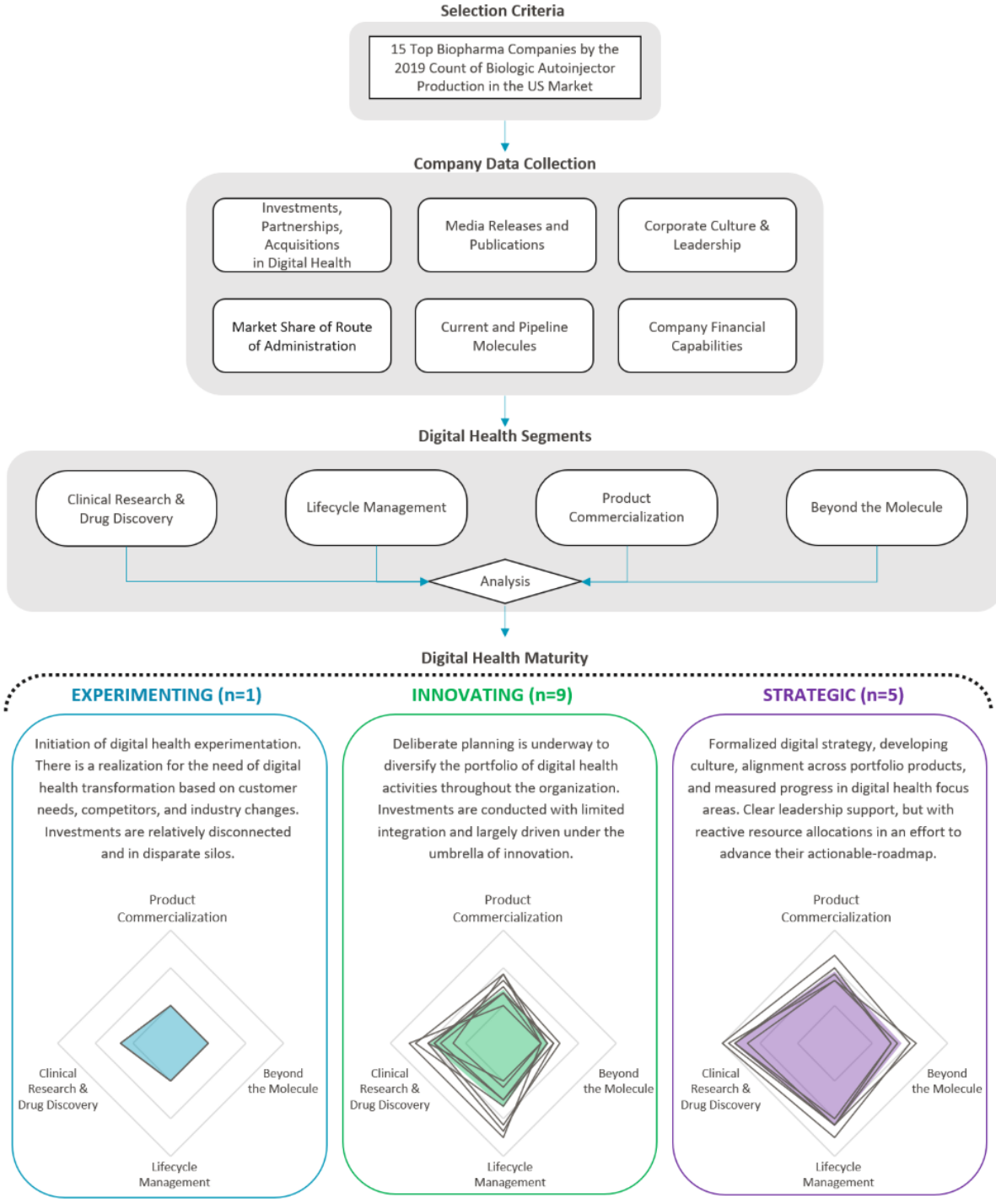
Analysis of Digital Health Maturity in Biopharma

We define digital health maturity as biopharma's organizational transformation by adopting digital health technologies, real-world evidence generation, digital-first leadership, and alignment of the product portfolio strategy. To assess the forces driving the next generation of autoinjectable devices, we quantified the digital health maturity of the top 15 biopharmaceutical companies in the US biologics autoinjector market. We conducted a detailed analysis of each company to evaluate their digital health activities across the following four segments: Clinical Research and Drug Discovery, Lifecycle Management, Product Commercialization, and Beyond the Molecule (Table 1). This framework for rating each biopharmaceutical company's maturity used qualitative and quantitative factors as described in Multimedia Appendix 1. The information from this maturity rating for each company is drawn from publicly available sources as of October 1, 2020, including US marketed and pipeline molecules, digital health–related strategic investments, partnerships and acquisitions, estimated spending committed to digital health endeavors, senior leadership's experience, and public statements addressing their digital health vision. A combination of public and private databases was used, including EvaluatePharma. To the degree that a biopharmaceutical company may have additional digital health initiatives in the abovementioned four segments that are not disclosed to the public, the company's maturity rating may be underestimated. The digital health maturity for each biopharmaceutical company is represented by a single number (between 0 and 1), which is the sum of each company's segmentation scores (0-20) normalized to the maximum value. The 15 companies were grouped into one of the following three categories: experimenting (bottom one-third), innovating (middle one-third), and strategic (top one-third), as shown in Figure 1.

Table 1. Four digital health segments used to evaluate biopharmaceutical company maturity.

Digital health segments	Definition
Clinical research and drug discovery	Process improvements in clinical research and drug development enabled by digital health to realize clinical benefits.
Lifecycle management	Continuous monitoring and improvement of the product or service through real-world evidence generation to meet the needs of the end user until product end of life.
Product commercialization	Digital health extensions of the molecule's capabilities outside of pure pharmacokinetic or pharmacodynamic impacts that increase the value of the molecule.
Beyond the molecule	Hardware or software solutions that provide therapeutic benefits independent of the molecule.

Figure 1. Framework for biopharma digital maturity assessment. This figure demonstrates biopharma's overall approach toward digital health based on a segmentation analysis covering clinical research and drug discovery, lifecycle management, product commercialization, and beyond the molecule. Companies were classified within three distinct digital health maturity categories. Radar charts show individual company ratings for each segment, and shaded regions represent the average for each category.

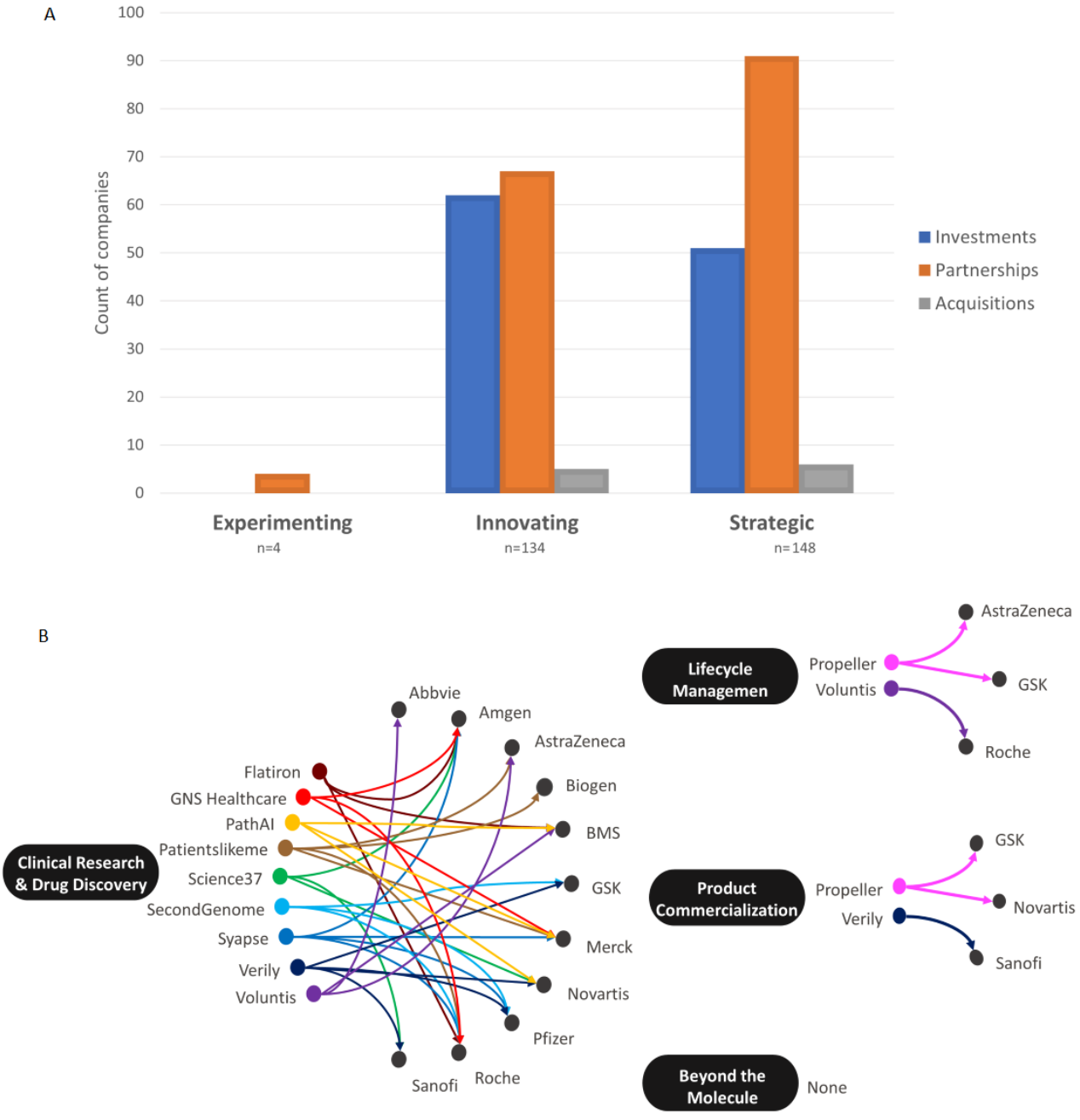


Key Themes of Biopharma's Digital Health Maturity

There is a clear differentiation among biopharma's digital health maturity, as seen by the radar charts in Figure 1. Each chart's center is skewed toward clinical research and drug discovery, demonstrating an increasing focus from companies who are adapting new technologies to their drug development processes. As we move from left to right across the categories, the area of the shaded region increases, representing expanding levels of digital health maturity across all four segments. It is interesting to note that where the experimenter focuses primarily on clinical research and drug discovery, innovators expand their interests to include lifecycle management with limited focus on product commercialization. At the same time, strategics show substantial activity in both product commercialization and beyond the molecule segments.

This analysis found 286 digital health companies working with biopharma. Considering the nature of biopharma's engagement with digital health, we classified the overall approach taken by companies in each category based upon investments, partnerships, and acquisitions (Figure 2A). Currently, biopharma is heavily vested in forming partnerships in digital health. However, there is a clear distinction with innovators and strategics who have a greater appetite for risk through investments and acquisitions. Across the board, biopharma's acquisitions have been limited so far, which may be due to a lack of perceived value or the complexities in integrating vastly different organizational cultures. Today, partnerships are the desired format for achieving digital health maturity. Figure 2B highlights prominent digital health companies defined as having a relationship with at least three or more of these biopharmaceutical companies. It is important to note that the four segments had varying trends, with a heavier focus on clinical research and drug discovery.

Figure 2. (A) Distribution of biopharma's digital health partnerships, investments, and acquisitions from 2010 to 2020. Investments included direct investments from biopharma or a subsidiary arm of the parent company. Partnerships are the preferred format for achieving digital health maturity. (B) Digital health companies with at least three different biopharma interactions (eg, investment, partnership, and acquisition) were plotted against the four digital health segments. Currently, clinical research and drug discovery digital health companies encapsulate the largest segment of investments or partnerships from biopharma. None: the analysis did not return any digital health companies focusing on the beyond the molecule segment, which met the minimum criteria of at least three different biopharma interactions.



Biopharma's core business is in clinical research and drug discovery, and as such, there is a primary focus on digital health efforts in this segment. Clinical trials are becoming increasingly complex and biologics are vastly more expensive to discover, so biopharma is using artificial intelligence to reduce attrition rates and research and development expenditure, and the vast data can accelerate the understanding of disease pathology and identify new drug targets and candidates [29-31]. Biopharma is turning toward digital health to not only improve the data collected from enrolled participants but also increase patient recruitment and retention (the largest cost driver of clinical trials) by engaging with patients through social media platforms or online health communities [32-35]. The premise is that by using digital health, they may shorten the time spent in clinical research while also amassing previously unattainable real-world data. Digital biomarkers will serve to generate novel data endpoints outside of traditional clinical environments and expand insights directly from the patient's home [8]. Biopharma will need to determine which digital biomarkers are valuable and how to integrate them into research. The overall digital health premise of data generation at the patient's home is highly attractive [36]. Our data reinforce this point as all companies regardless of their maturity category have applied digital health to their clinical research and drug discovery [37].

The second largest segment is lifecycle management where its implementation varies across companies as influenced by their digital health maturity. Its most basic manifestation consists of packaging or a companion app, while higher sophistication levels have been demonstrated through connected DDCPs [38]. These products can be considered to provide support for biopharma's drugs and often fall in line with their therapeutic portfolios, generate real-world evidence, and aid in gathering novel data sets to differentiate and extend the longevity of their molecules. One of the key themes was user-focused mobile apps to empower patient disease self-management; however, a high churn rate remains a strong barrier to achieving the desired health outcomes [39]. Our findings clearly demonstrate that innovating and strategic companies have made considerable inroads with the application of digital health to lifecycle management.

As companies continue to commercialize their drug products, few are proactively designing these products with integrated digital capabilities. The minimum design for these devices is Bluetooth connectivity to a patient-facing mobile app. This connected DDCP can act as an adherence measure, allowing patients to keep a record of their medication use and share their data with their providers [40,41]. Some companies have taken a further step by creating entire platforms for their devices, including daily predictive forecasts and integrated and streamlined communication with health care providers and support programs. Expanding beyond this includes integrated sensors and apps enabling drugs to aid in chronic disease management, which can collect a variety of data from general adherence to pharmacokinetic and pharmacodynamic data that could better inform patients and their providers about usage behaviors to optimize drug adherence and treatment [9,42,43]. The next step will be fully connected DDCPs dispensed from the pharmacy, which has been seen with bioingestible sensors in oral medications and connected inhalers [44,45]. Both innovators and strategics have achieved limited product commercialization and have not yet fully explored how user error could impact collected data in chronic disease management [46].

Of all the segments evaluated for biopharma's digital health maturity, beyond the molecule is the least explored, demonstrating the reservations toward digital therapeutics. Multiple companies in the innovating and strategic groups have partnered with a digital therapeutic company. As seen in Figure 2B, no digital health company has succeeded in attracting multiple biopharma interests in beyond the molecule solutions compared to those focused on other segments. Digital therapeutic companies' narrow therapeutic focuses may not align them across multiple biopharma pipelines at this time. Other areas of emerging focus in the beyond the molecule segment include gamification technology and virtual reality that could offer novel therapeutic treatments [47-50]. The digital therapeutics space is an area that will blossom; however, the push to embrace a beyond the molecule business model is in its infancy today and future growth is largely expected to be driven by strategics and a few innovators [51].

Biopharma's internal cultural dynamics can influence an organization's digital health maturity. To better understand internal leadership culture, individual digital health leadership industry backgrounds were aggregated across the previous 15 years based on their disclosed roles on a professional social media platform (ie, LinkedIn). As biopharma's maturity increased, leaders had more diversified backgrounds and companies relied less on promotion from within the biopharmaceutical industry (Table 2). This clear correlation may attribute low digital maturity to a lack of outside novel perspective. One of the core limiting factors encumbering biopharma is the lack of personnel with formal education and training in digital health. Relatively few programs are currently focusing on digital health training, as seen by the lack of standard practices and education [52]. As such, there is a definitive digital health skills shortage across biopharma, and this is demonstrated by the largest sector of personnel being internally promoted to digital health–focused divisions, regardless of the company's maturity level.

Table 2. Digital health leadership backgrounds.

Industry[a], (n=235)	Experimenting	Innovative	Strategic
Biopharma	76%	53%	44%
Health care	0%	10%	12%
Marketing	0%	2%	4%
Medical devices	0%	3%	2%
Consulting	3%	13%	12%
Education	14%	6%	2%
Finance	0%	1%	2%
Information technology	7%	4%	6%
Others[b]	0%	8%	16%

[a]Internal biopharma digital health leadership backgrounds have been evaluated across the three maturity categories. The sample size includes 235 digital health employees that hold an executive, head, or director level position.

[b]"Others" include: Retail, Hospitality, Food & Beverage, Telecom, Media, VC, Government, Entertainment, Utilities, Staffing, Insurance, Renewable, Travel, Apparel, Cosmetics, Research, Consumer Goods, Law, and Farming.

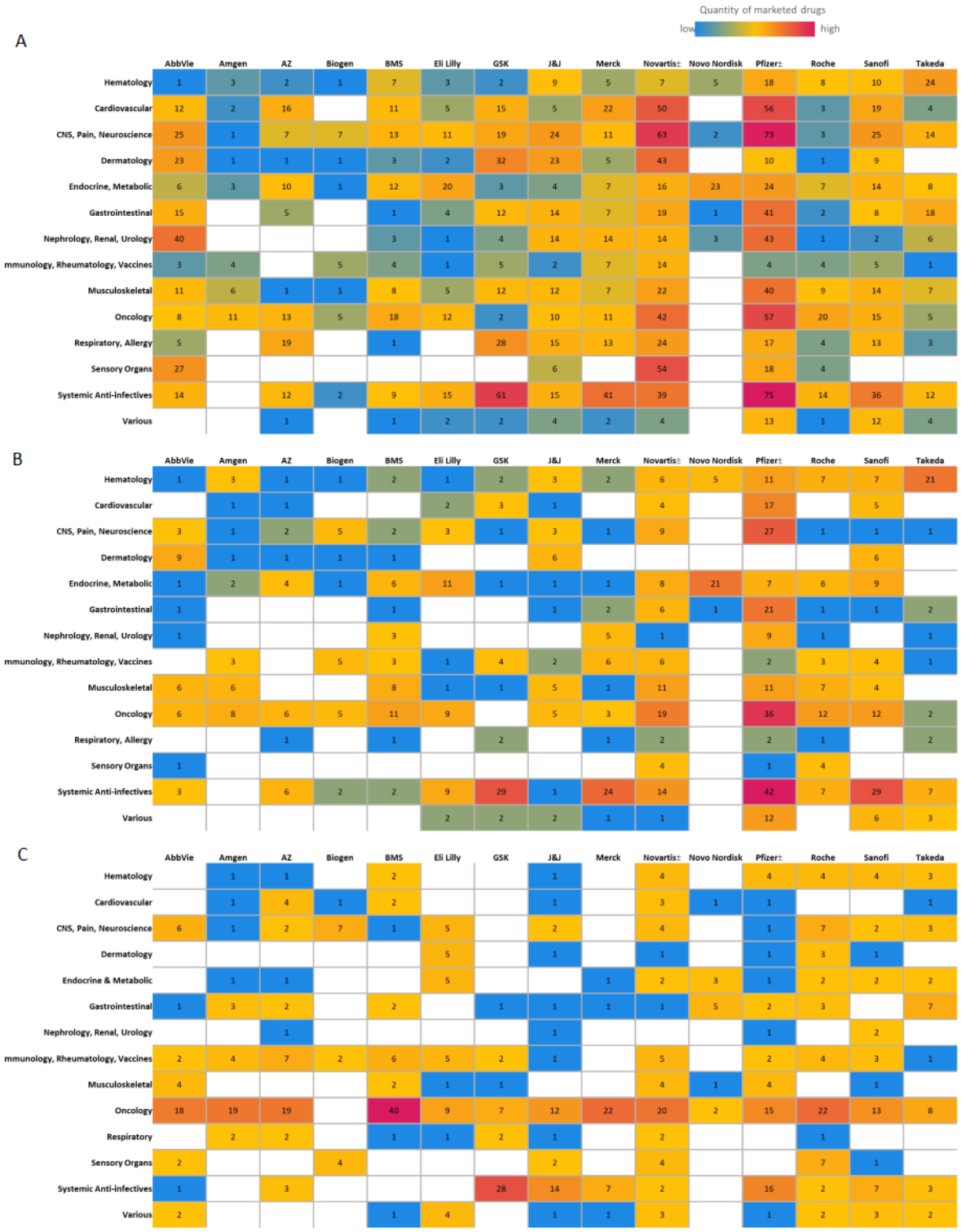
Evolving Biopharma Pipeline Driving Innovation in Autoinjectable Devices

Many digital health forays have been focused on clinical research activities. Figure 3A-C presents heat maps showing the number of drug molecules by biopharmaceutical companies and therapeutic categories based on their US primary indication for (1) currently marketed molecules, (2) currently marketed injectables, and (3) active injectable drugs in the pipeline. When comparing current marketed injectables (Figure 3B) to pipeline injectables (Figure 3C), there is a noticeable shift across the majority of companies toward autoimmune conditions and oncology.

In the autoimmune category, the major biopharmaceutical companies have increased their deal making with respect to acquiring products to fill their pipelines. Both in-licensing and out-licensing activities have shown notable increases in the last few years (2013-2018) [53]. Biopharmaceutical companies licensing-in products are paying more than twice as much for new autoimmune products in the recent period (ie, 2013-2017) than they did in the previous 5-year period (ie, 2008-2012) [51].

Oncology is the leading therapeutic category in the injectable pipeline. Oncology deal-making saw an increase of 142% in the period of 2013 to 2017, with 643 deals compared with 266 deals in the period of 2008 to 2012 [53]. Oncology is the major therapeutic focus in the injectable pipeline for 12 of the 15 companies evaluated in this study. Interestingly, research is being focused on subcutaneous delivery of oncology products, with trastuzumab researched extensively [37,53-55]. With the discovery of more biologic therapeutic agents, we see more cancer patients being treated at home rather than in controlled inpatient settings. Moving forward, the administration of advanced biotechnology-derived agents will be more prevalent in the home environment. Oncology specifically offers tremendous market potential for drug products engaged with digital health technologies to address various unmet needs for patient care [56]. The expanding oncology pipeline can be combined with novel research approaches using site-less trial designs to study and deliver effective therapies to otherwise high-value therapeutic markets [57].

Figure 3. Biopharma portfolios and pipeline by therapeutic areas. Heatmaps showing (A) marketed molecules across all routes of administration, (B) marketed injectable molecules, and (C) phase I to III pipeline molecules proposed for injectable administration. Biopharma pipeline assessment for the US market was conducted via the EvaluatePharma database as of October 1, 2020. Therapeutic areas defined in this paper were standardized across each biopharmaceutical company. For Novartis and Pfizer, generic manufacturing has been included in the count.

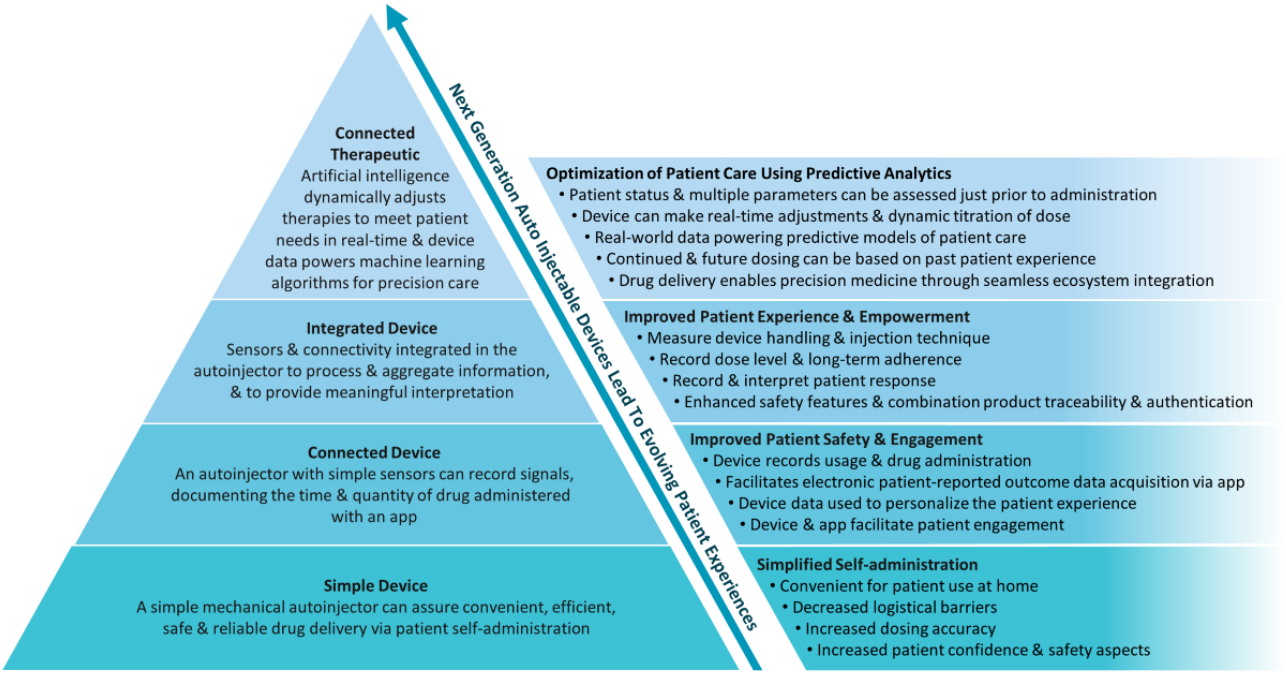


Framework for Next-Generation Connected Autoinjectable Devices

The framework for a connected autoinjectable device is multifaceted but resides in addressing today's unmet need for patient self-administration in the home. Figure 4 highlights the autoinjector's evolution trajectory from a simple device to connected therapeutics. In its most basic form, the autoinjector is designed to facilitate patient self-administration via a simple process, while retaining patient convenience and safety. This exchange does not elicit measurable objective outcomes for stakeholders, such as patient adherence, injection technique, and patient-reported outcomes. Moore law, combined with

advancements in digital health applications, is now enabling the development of connected or integrated autoinjectors to varying utilization levels [58]. These sensors and communication modules not only elevate the safety aspects of the device (eg, temperature monitoring, authentication tags, recalls, and tampering alerts), but also track measurable patient outcomes (eg, dose-level adherence, injection technique, side-effects, and therapeutic outcomes) tethered to a mobile app or hub for data acquisition, transmission, and analysis. These connected and integrated autoinjectors can now enhance patient engagement and guide therapeutic decision making with objective data outputs. A connected therapeutics product then shifts the data value away from the clinic and to the home, when high-resolution objective data outputs from connected or integrated autoinjectors are captured to power machine learning predictive models to reliably inform real-time care decision making. The autoinjector then transitions to become the focal point of decentralized precision care for many chronic conditions, enabling artificial intelligence disease management systems to track or predict patient outcomes (eg, therapeutic outcomes, major events, experiences, and side effects) both individually and at the population level. Overall, connected therapeutics is the highest evolution of the connected autoinjector, which is represented by the pyramid peak in Figure 4.

Figure 4. Framework for next-generation connected autoinjectable devices. This framework for next-generation connected autoinjectors demonstrates the technological hierarchy of design (left) that transitions the simple device to a connected therapeutic. Through this design evolution, the autoinjector shifts to become the focal point of decentralized precision care for many chronic conditions, powering artificial intelligence disease management systems that impact overall patient care (right).



As previously identified, biopharma will be the leader in championing this paradigm shift toward connected therapeutics, but internal culture and leadership will likely dictate adoption over the coming years. For autoinjectors to no longer be viewed as simple devices facilitating drug administration, but as an enabling technology for decentralized precision care, a shift in mindset is required. To create this shift, biopharma will need to foster an innovation culture, achieve digital health workforce diversity, and leverage partnerships outside the biopharmaceutical industry. We have demonstrated a considerable discord in digital health leadership across all three maturity categories in our analysis. Most evident are the strategics with the most diversified leadership, encapsulating outside novel perspectives and demonstrating the greatest appetite for external partnerships across the four digital health segments, which sets them apart from experimenters and innovators. As such, more mature companies will likely extend the value proposition of connected autoinjectors and expand to connected therapeutics through funding research and engaging providers and payors for market-shaping strategies. In comparison, it is expected that experimenters and innovators will focus on connected autoinjectors for their current products' lifecycle management as they determine how to achieve broader market differentiation until their leadership envisions a digital health strategy.

An evidence-based approach, utilizing interdisciplinary teams of clinical, engineering, economic, and behavioral science experts, will be critical for demonstrating the feasibility of next-generation connected autoinjectable devices [58]. Our evidence generation value proposition is however not necessarily novel [59]. The connected inhaler space has shown multiple successes and generated considerable evidence as a model to follow, but the application of the body of evidence is narrow owing to the few therapeutic categories (eg, asthma and chronic obstructive pulmonary disease) that inhalers address [60]. In comparison, the autoinjector market is much more differentiated across a whole spectrum of conditions, demanding an evidence generation process that is specific to each patient population, the disease condition being addressed, and the pharmacotherapy

[61]. This will be a significant endeavor as our analysis has demonstrated a growing pipeline of therapies focused on chronic conditions using injectable molecules, in addition to the likelihood of oncology becoming a growing market for home self-administration [62]. While a focus on adherence is currently the simplest business case to encourage evolution of the autoinjector toward connected therapeutics, biopharma will also need to demonstrate improvements in clinical outcomes via providers, increased economic efficiencies via payors, and patient satisfaction and usability via sustained levels of engagement [63]. Additional concerns will be on how connected therapeutics will integrate within a broader health care ecosystem, embracing remote diagnostics and digital therapeutics–augmented treatments, in order to enrich actionable data sets and reduce data silos that have historically led to a poor uptake of digital health interventions [64].

Conclusion

Next-generation autoinjectable devices will play an important role in implementing biopharma's digital health approach to the biologics market. Our analysis demonstrated considerable biopharma maturation differences with digital health. In the coming decade, biopharma will need to design a strategic and methodological pathway to embed digital health as a key corporate cultural aspect in order to succeed. Utilizing digital health, connected therapeutics will allow biopharma to achieve a closer relationship with patients in the home and with providers, as our framework establishes. With autoinjectable devices enabling home self-administration and connected therapeutics powering the delivery of precision care at home, biopharma will need to drive innovation in autoinjectors to achieve the full potential of biologics.

Acknowledgments

The authors acknowledge Jamie Gould, Michael McGowan, Zora Lin, Marcus Clinch, and Simeon Simone for their support in refining the intellectual content, data analysis, and interpretation.

Authors' Contributions

All authors contributed to the conception of the viewpoint, including analysis and interpretation of the data, and the drafting and refining of this work.

Conflicts of Interest

TDA has served as a consultant or advisor for Otsuka Pharmaceuticals, Teva, and Eli Lilly. He is on Otsuka Pharmaceuticals speaker's bureau. He is also an advisor for HealthXL and The Digital Therapeutics Alliance. The other authors declare no competing interests.

Multimedia Appendix 1
Qualitative and quantitative factors used to establish the maturity rating for each biopharmaceutical company.
[DOCX File , 29 KB - mhealth_v9i3e25406_app1.docx]

References

1. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB. Digital health: a path to validation. NPJ Digit Med 2019;2:38 [FREE Full text] [doi: 10.1038/s41746-019-0111-3] [Medline: 31304384]
2. Snowden S. Digital Health: A Framework for Healthcare Transformation. Healthcare Information and Management System Society. 2020. URL: https://www.himss.org/resources/digital-health-framework-healthcare-transformation-white-paper [accessed 2020-09-21]
3. What is Digital Health? FDA. URL: https://www.fda.gov/medical-devices/digital-health-center-excellence/what-digital-health [accessed 2020-09-21]
4. Turakhia MP, Desai SA, Harrington RA. The Outlook of Digital Health for Cardiovascular Medicine: Challenges but Also Extraordinary Opportunities. JAMA Cardiol 2016 Oct 01;1(7):743-744. [doi: 10.1001/jamacardio.2016.2661] [Medline: 27580275]
5. 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]
6. Jones GB, Collins DS, Harrison MW, Thyagarajapuram NR, Wright JM. Subcutaneous drug delivery: An evolving enterprise. Sci Transl Med 2017 Aug 30;9(405):eaaf9166. [doi: 10.1126/scitranslmed.aaf9166] [Medline: 28855399]
7. Harvard Pilgrim Signs Outcomes-Based Contract with Amgen for Enbrel. Harvard Pilgrim Health Care. 2017 Feb 22. URL: https://www.harvardpilgrim.org/public/news-detail?nt=HPH_News_C&nid=1471912468296 [accessed 2020-09-15]
8. Coravos A, Khozin S, Mandl KD. Erratum: Author Correction: Developing and adopting safe and effective digital biomarkers to improve patient outcomes. NPJ Digit Med 2019 May 10;2(1):40 [FREE Full text] [doi: 10.1038/s41746-019-0119-8] [Medline: 31304386]

9. Bittner B, Schmit Chiesi C, Kharawala S, Kaur G, Schmidt J. Connected drug delivery devices to complement drug treatments: potential to facilitate disease management in home setting. Med Devices (Auckl) 2019;12:101-127 [FREE Full text] [doi: 10.2147/MDER.S198943] [Medline: 30881151]
10. McGowan MM. 30 Years in the Making – an Inside Perspective on the Emergence of Autoinjectors. ONdrugDelivery Magazine. 2019 Oct. URL: https://www.ondrugdelivery.com/30-years-in-the-making-an-inside-perspective-on-the-emergence-of-autoinjectors/ [accessed 2020-09-15]
11. Compilation of CDER New Molecular Entity (NME) Drug and New Biologic Approvals. FDA. URL: https://www.fda.gov/media/135307/download [accessed 2020-09-15]
12. Thakur K, Biberger A, Handrich A, Rezk MF. Patient Perceptions and Preferences of Two Etanercept Autoinjectors for Rheumatoid Arthritis: Findings from a Patient Survey in Europe. Rheumatol Ther 2016 Dec 5;3(2):245-256 [FREE Full text] [doi: 10.1007/s40744-016-0048-9] [Medline: 27817152]
13. Thakur K, Biberger A, Handrich A, Rezk MF. Perceptions and Preferences of Two Etanercept Autoinjectors for Rheumatoid Arthritis: A New European Union-Approved Etanercept Biosimilar (Benepali) Versus Etanercept (Enbrel) - Findings from a Nurse Survey in Europe. Rheumatol Ther 2016 Jun 23;3(1):77-89 [FREE Full text] [doi: 10.1007/s40744-016-0035-1] [Medline: 27747519]
14. Bayas A. Improving adherence to injectable disease-modifying drugs in multiple sclerosis. Expert Opin Drug Deliv 2013 Mar 23;10(3):285-287. [doi: 10.1517/17425247.2013.763793] [Medline: 23339371]
15. Conway CM, Kelechi TJ. Digital Health for Medication Adherence in Adult Diabetes or Hypertension: An Integrative Review. JMIR Diabetes 2017 Aug 16;2(2):e20 [FREE Full text] [doi: 10.2196/diabetes.8030] [Medline: 30291093]
16. Alipanah N, Jarlsberg L, Miller C, Linh NN, Falzon D, Jaramillo E, et al. Adherence interventions and outcomes of tuberculosis treatment: A systematic review and meta-analysis of trials and observational studies. PLoS Med 2018 Jul 3;15(7):e1002595 [FREE Full text] [doi: 10.1371/journal.pmed.1002595] [Medline: 29969463]
17. Checchi KD, Huybrechts KF, Avorn J, Kesselheim AS. Electronic medication packaging devices and medication adherence: a systematic review. JAMA 2014 Sep 24;312(12):1237-1247 [FREE Full text] [doi: 10.1001/jama.2014.10059] [Medline: 25247520]
18. Kini V, Ho PM. Interventions to Improve Medication Adherence: A Review. JAMA 2018 Dec 18;320(23):2461-2473. [doi: 10.1001/jama.2018.19271] [Medline: 30561486]
19. Watanabe JH, McInnis T, Hirsch JD. Cost of Prescription Drug-Related Morbidity and Mortality. Ann Pharmacother 2018 Sep 26;52(9):829-837. [doi: 10.1177/1060028018765159] [Medline: 29577766]
20. DiMatteo MR. Variations in patients' adherence to medical recommendations: a quantitative review of 50 years of research. Med Care 2004 Mar;42(3):200-209. [doi: 10.1097/01.mlr.0000114908.90348.f9] [Medline: 15076819]
21. Neiman AB, Ruppar T, Ho M, Garber L, Weidle PJ, Hong Y, et al. CDC Grand Rounds: Improving Medication Adherence for Chronic Disease Management - Innovations and Opportunities. MMWR Morb Mortal Wkly Rep 2017 Nov 17;66(45):1248-1251 [FREE Full text] [doi: 10.15585/mmwr.mm6645a2] [Medline: 29145353]
22. Grijalva CG, Chung CP, Arbogast PG, Stein CM, Mitchel EF, Griffin MR. Assessment of adherence to and persistence on disease-modifying antirheumatic drugs (DMARDs) in patients with rheumatoid arthritis. Med Care 2007 Oct;45(10 Supl 2):S66-S76. [doi: 10.1097/MLR.0b013e318041384c] [Medline: 17909386]
23. Yazici Y, Krasnokutsky S, Barnes JP, Hines PL, Wang J, Rosenblatt L. Changing patterns of tumor necrosis factor inhibitor use in 9074 patients with rheumatoid arthritis. J Rheumatol 2009 May 30;36(5):907-913. [doi: 10.3899/jrheum.080592] [Medline: 19332636]
24. Harnett J, Wiederkehr D, Gerber R, Gruben D, Bourret J, Koenig A. Primary Nonadherence, Associated Clinical Outcomes, and Health Care Resource Use Among Patients with Rheumatoid Arthritis Prescribed Treatment with Injectable Biologic Disease-Modifying Antirheumatic Drugs. J Manag Care Spec Pharm 2016 Mar;22(3):209-218. [doi: 10.18553/jmcp.2016.22.3.209] [Medline: 27003550]
25. Bluett J, Morgan C, Thurston L, Plant D, Hyrich KL, Morgan AW, BRAGGSS. Impact of inadequate adherence on response to subcutaneously administered anti-tumour necrosis factor drugs: results from the Biologics in Rheumatoid Arthritis Genetics and Genomics Study Syndicate cohort. Rheumatology (Oxford) 2015 Mar;54(3):494-499 [FREE Full text] [doi: 10.1093/rheumatology/keu358] [Medline: 25213131]
26. Ellery T, Hansen N, editors. Biologics and LCM. In: Pharmaceutical Lifecycle Management: Making the Most of Each and Every Brand. Hoboken, NJ, USA: Wiley; 2012:207-216.
27. Sangave NA, Aungst TD, Patel DK. Smart Connected Insulin Pens, Caps, and Attachments: A Review of the Future of Diabetes Technology. Diabetes Spectr 2019 Nov 14;32(4):378-384 [FREE Full text] [doi: 10.2337/ds18-0069] [Medline: 31798296]
28. Park S, Garcia-Palacios J, Cohen A, Varga Z. From treatment to prevention: The evolution of digital healthcare. Nature. URL: https://media.nature.com/original/magazine-assets/d42473-019-00274-6/d42473-019-00274-6.pdf [accessed 2020-09-15]
29. Digital Health Center of Excellence. FDA. URL: https://www.fda.gov/medical-devices/digital-health-center-excellence [accessed 2020-09-15]

30. Moore TJ, Zhang H, Anderson G, Alexander GC. Estimated Costs of Pivotal Trials for Novel Therapeutic Agents Approved by the US Food and Drug Administration, 2015-2016. JAMA Intern Med 2018 Nov 01;178(11):1451-1457 [FREE Full text] [doi: 10.1001/jamainternmed.2018.3931] [Medline: 30264133]
31. Mak K, Pichika MR. Artificial intelligence in drug development: present status and future prospects. Drug Discov Today 2019 Mar;24(3):773-780. [doi: 10.1016/j.drudis.2018.11.014] [Medline: 30472429]
32. Schuhmacher A, Gatto A, Hinder M, Kuss M, Gassmann O. The upside of being a digital pharma player. Drug Discov Today 2020 Sep;25(9):1569-1574. [doi: 10.1016/j.drudis.2020.06.002] [Medline: 32554063]
33. Inan OT, Tenaerts P, Prindiville SA, Reynolds HR, Dizon DS, Cooper-Arnold K, et al. Digitizing clinical trials. NPJ Digit Med 2020 Jul 31;3(1):101 [FREE Full text] [doi: 10.1038/s41746-020-0302-y] [Medline: 32821856]
34. Jessup DL, Glover Iv MK, Daye D, Banzi L, Jones P, Choy G, et al. Implementation of Digital Awareness Strategies to Engage Patients and Providers in a Lung Cancer Screening Program: Retrospective Study. J Med Internet Res 2018 Feb 15;20(2):e52 [FREE Full text] [doi: 10.2196/jmir.8932] [Medline: 29449199]
35. Costello RE, Anand A, Jameson Evans M, Dixon WG. Associations Between Engagement With an Online Health Community and Changes in Patient Activation and Health Care Utilization: Longitudinal Web-Based Survey. J Med Internet Res 2019 Aug 29;21(8):e13477 [FREE Full text] [doi: 10.2196/13477] [Medline: 31469082]
36. Marra C, Chen JL, Coravos A, Stern AD. Quantifying the use of connected digital products in clinical research. NPJ Digit Med 2020 Apr 03;3(1):50 [FREE Full text] [doi: 10.1038/s41746-020-0259-x] [Medline: 32285011]
37. Kourtis LC, Regele OB, Wright JM, Jones GB. Digital biomarkers for Alzheimer's disease: the mobile/ wearable devices opportunity. NPJ Digit Med 2019;2 [FREE Full text] [doi: 10.1038/s41746-019-0084-2] [Medline: 31119198]
38. Bittner B, Richter W, Schmidt J. Subcutaneous Administration of Biotherapeutics: An Overview of Current Challenges and Opportunities. BioDrugs 2018 Oct 24;32(5):425-440 [FREE Full text] [doi: 10.1007/s40259-018-0295-0] [Medline: 30043229]
39. Huang Z, Tan E, Lum E, Sloot P, Boehm BO, Car J. A Smartphone App to Improve Medication Adherence in Patients With Type 2 Diabetes in Asia: Feasibility Randomized Controlled Trial. JMIR Mhealth Uhealth 2019 Sep 12;7(9):e14914 [FREE Full text] [doi: 10.2196/14914] [Medline: 31516127]
40. Baumel A, Muench F, Edan S, Kane JM. Objective User Engagement With Mental Health Apps: Systematic Search and Panel-Based Usage Analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
41. Harbison R, Hecht M, MacLeod J. Building a Data-Driven Multiple Daily Insulin Therapy Model Using Smart Insulin Pens. J Diabetes Sci Technol 2020 Aug 24:1932296820951225. [doi: 10.1177/1932296820951225] [Medline: 32830521]
42. Adolfsson P, Hartvig NV, Kaas A, Møller JB, Hellman J. Increased Time in Range and Fewer Missed Bolus Injections After Introduction of a Smart Connected Insulin Pen. Diabetes Technol Ther 2020 Oct 01;22(10):709-718 [FREE Full text] [doi: 10.1089/dia.2019.0411] [Medline: 32003590]
43. Limmroth V, Hechenbichler K, Müller C, Schürks M. Assessment of Medication Adherence Using a Medical App Among Patients With Multiple Sclerosis Treated With Interferon Beta-1b: Pilot Digital Observational Study (PROmyBETAapp). J Med Internet Res 2019 Jul 29;21(7):e14373 [FREE Full text] [doi: 10.2196/14373] [Medline: 31359863]
44. Domańska B, Vansant S, Mountian I. A Companion App to Support Rheumatology Patients Treated with Certolizumab Pegol: Results From a Usability Study. JMIR Form Res 2020 Jul 27;4(7):e17373 [FREE Full text] [doi: 10.2196/17373] [Medline: 32716310]
45. Mehta PP. Dry powder inhalers: a concise summary of the electronic monitoring devices. Ther Deliv 2021 Jan;12(1):1-6 [FREE Full text] [doi: 10.4155/tde-2020-0091] [Medline: 32873214]
46. Gunasekeran DV. Technology and chronic disease management. The Lancet Diabetes & Endocrinology 2018 Feb;6(2):91. [doi: 10.1016/S2213-8587(17)30441-2] [Medline: 29413164]
47. No authors listed. Aripiprazole with digital ingestion tracking (Abilify MyCite). Med Lett Drugs Ther 2019 Jan 28;61(1564):15-16. [Medline: 30856161]
48. Floryan M, Chow PI, Schueller SM, Ritterband LM. The Model of Gamification Principles for Digital Health Interventions: Evaluation of Validity and Potential Utility. J Med Internet Res 2020 Jun 10;22(6):e16506 [FREE Full text] [doi: 10.2196/16506] [Medline: 32519965]
49. van de Weijer SC, Duits AA, Bloem BR, de Vries NM, Kessels RP, Köhler S, et al. Feasibility of a Cognitive Training Game in Parkinson's Disease: The Randomized Parkin'Play Study. Eur Neurol 2020 Aug 5;83(4):426-432 [FREE Full text] [doi: 10.1159/000509685] [Medline: 32756067]
50. Gates NJ, Vernooij RW, Di Nisio M, Karim S, March E, Martínez G, et al. Computerised cognitive training for preventing dementia in people with mild cognitive impairment. Cochrane Database Syst Rev 2019 Mar 13;3:CD012279 [FREE Full text] [doi: 10.1002/14651858.CD012279.pub2] [Medline: 30864747]
51. Sverdlov O, van Dam J, Hannesdottir K, Thornton-Wells T. Digital Therapeutics: An Integral Component of Digital Innovation in Drug Development. Clin Pharmacol Ther 2018 Jul 23;104(1):72-80. [doi: 10.1002/cpt.1036] [Medline: 29377057]

52. Aungst TD, Patel R. Integrating Digital Health into the Curriculum-Considerations on the Current Landscape and Future Developments. J Med Educ Curric Dev 2020 Jan 20;7:2382120519901275 [FREE Full text] [doi: 10.1177/2382120519901275] [Medline: 32010795]
53. Micklus A, Muntner S. Biopharma deal-making in 2017. Nat Rev Drug Discov 2018 Feb 01;17(2):91-92. [doi: 10.1038/nrd.2018.1] [Medline: 29386610]
54. Hopson S, Casebeer A, Stemkowski S, Antol DD, Tao Z, Howe A, et al. Does site-of-care for oncology infusion therapy influence treatment patterns, cost, and quality in the United States? J Med Econ 2018 Feb 17;21(2):152-162. [doi: 10.1080/13696998.2017.1384736] [Medline: 28945163]
55. Jackisch C, Stroyakovskiy D, Pivot X, Ahn JS, Melichar B, Chen S, et al. Subcutaneous vs Intravenous Trastuzumab for Patients With ERBB2-Positive Early Breast Cancer: Final Analysis of the HannaH Phase 3 Randomized Clinical Trial. JAMA Oncol 2019 May 01;5(5):e190339 [FREE Full text] [doi: 10.1001/jamaoncol.2019.0339] [Medline: 30998824]
56. Garg S, Williams NL, Ip A, Dicker AP. Clinical Integration of Digital Solutions in Health Care: An Overview of the Current Landscape of Digital Technologies in Cancer Care. JCO Clin Cancer Inform 2018 Dec;2:1-9 [FREE Full text] [doi: 10.1200/CCI.17.00159] [Medline: 30652580]
57. Persky S. A Virtual Home for the Virtual Clinical Trial. J Med Internet Res 2020 Jan 03;22(1):e15582 [FREE Full text] [doi: 10.2196/15582] [Medline: 31899455]
58. Rafiei RR. Macro Trends Accelerating the Adoption of Connectivity in Drug Delivery. ONdrugDelivery Magazine. 2019. URL: https://www.ondrugdelivery.com/macro-trends-accelerating-the-adoption-of-connectivity-in-drug-delivery/ [accessed 2020-09-15]
59. Guo C, Ashrafian H, Ghafur S, Fontana G, Gardner C, Prime M. Challenges for the evaluation of digital health solutions-A call for innovative evidence generation approaches. NPJ Digit Med 2020 Aug 27;3(1):110 [FREE Full text] [doi: 10.1038/s41746-020-00314-2] [Medline: 32904379]
60. 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]
61. Morton RW, Elphick HE, Rigby AS, Daw WJ, King DA, Smith LJ, et al. STAAR: a randomised controlled trial of electronic adherence monitoring with reminder alarms and feedback to improve clinical outcomes for children with asthma. Thorax 2017 Apr 04;72(4):347-354. [doi: 10.1136/thoraxjnl-2015-208171] [Medline: 27815524]
62. Boeri M, Szegvari B, Hauber B, Mange B, Mountian I, Schiff M, et al. From drug-delivery device to disease management tool: a study of preferences for enhanced features in next-generation self-injection devices. Patient Prefer Adherence 2019 Jul;13:1093-1110 [FREE Full text] [doi: 10.2147/PPA.S203775] [Medline: 31371927]
63. Pratap A, Neto EC, Snyder P, Stepnowsky C, Elhadad N, Grant D, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. NPJ Digit Med 2020 Feb 17;3(1):21 [FREE Full text] [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
64. Shull JG. Digital Health and the State of Interoperable Electronic Health Records. JMIR Med Inform 2019 Nov 01;7(4):e12712 [FREE Full text] [doi: 10.2196/12712] [Medline: 31682583]

Abbreviations

DDCP: drug device combination product

Edited by L Buis; submitted 31.10.20; peer-reviewed by R Patel, D Gunasekeran; comments to author 17.11.20; revised version received 23.11.20; accepted 13.02.21; published 18.03.21.

Please cite as:
Rafiei R, Williams C, Jiang J, Aungst TD, Durrer M, Tran D, Howald R
Digital Health Integration Assessment and Maturity of the United States Biopharmaceutical Industry: Forces Driving the Next Generation of Connected Autoinjectable Devices
JMIR Mhealth Uhealth 2021;9(3):e25406
URL: https://mhealth.jmir.org/2021/3/e25406
doi:10.2196/25406
PMID:33621188

Review

Digital Technology Interventions for Risk Factor Modification in Patients With Cardiovascular Disease: Systematic Review and Meta-analysis

Adewale Samuel Akinosun[1], BSc, MPH; Rob Polson[2], MSc; Yohanca Diaz - Skeete[3], BSc, MSc; Johannes Hendrikus De Kock[1], PhD; Lucia Carragher[3], PhD; Stephen Leslie[4], MBChB, PhD; Mark Grindle[1], MLitt, PhD; Trish Gorely[1], BPhEd, MEd, PhD

[1]Department of Nursing and Midwifery, Institute of Health Research and Innovation, Centre for Health Science, University of the Highlands and Islands, Inverness, United Kingdom

[2]Highland Health Sciences Library, Centre for Health Science, University of the Highlands and Islands, Inverness, United Kingdom

[3]School of Health and Science, Dundalk Institute of Technology, Dundalk, Ireland

[4]Cardiology Unit, Raigmore Hospital, NHS Highlands, Inverness, United Kingdom

Corresponding Author:
Adewale Samuel Akinosun, BSc, MPH
Department of Nursing and Midwifery, Institute of Health Research and Innovation
Centre for Health Science
University of the Highlands and Islands
Inverness
United Kingdom
Phone: 44 07727867360
Email: 18026171@uhi.ac.uk

Abstract

Background: Approximately 50% of cardiovascular disease (CVD) cases are attributable to lifestyle risk factors. Despite widespread education, personal knowledge, and efficacy, many individuals fail to adequately modify these risk factors, even after a cardiovascular event. Digital technology interventions have been suggested as a viable equivalent and potential alternative to conventional cardiac rehabilitation care centers. However, little is known about the clinical effectiveness of these technologies in bringing about behavioral changes in patients with CVD at an individual level.

Objective: The aim of this study is to identify and measure the effectiveness of digital technology (eg, mobile phones, the internet, software applications, wearables, etc) interventions in randomized controlled trials (RCTs) and determine which behavior change constructs are effective at achieving risk factor modification in patients with CVD.

Methods: This study is a systematic review and meta-analysis of RCTs designed according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analysis) statement standard. Mixed data from studies extracted from selected research databases and filtered for RCTs only were analyzed using quantitative methods. Outcome hypothesis testing was set at 95% CI and P=.05 for statistical significance.

Results: Digital interventions were delivered using devices such as cell phones, smartphones, personal computers, and wearables coupled with technologies such as the internet, SMS, software applications, and mobile sensors. Behavioral change constructs such as cognition, follow-up, goal setting, record keeping, perceived benefit, persuasion, socialization, personalization, rewards and incentives, support, and self-management were used. The meta-analyzed effect estimates (mean difference [MD]; standard mean difference [SMD]; and risk ratio [RR]) calculated for outcomes showed benefits in total cholesterol SMD at −0.29 [−0.44, −0.15], P<.001; high-density lipoprotein SMD at –0.09 [–0.19, 0.00], P=.05; low-density lipoprotein SMD at −0.18 [−0.33, −0.04], P=.01; physical activity (PA) SMD at 0.23 [0.11, 0.36], P<.001; physical inactivity (sedentary) RR at 0.54 [0.39, 0.75], P<.001; and diet (food intake) RR at 0.79 [0.66, 0.94], P=.007. Initial effect estimates showed no significant benefit in body mass index (BMI) MD at −0.37 [−1.20, 0.46], P=.38; diastolic blood pressure (BP) SMD at −0.06 [−0.20, 0.08], P=.43; systolic BP SMD at −0.03 [−0.18, 0.13], P=.74; Hemoglobin A_{1C} blood sugar (HbA_{1c}) RR at 1.04 [0.40, 2.70], P=.94; alcohol intake SMD at −0.16 [−1.43, 1.10], P=.80; smoking RR at 0.87 [0.67, 1.13], P=.30; and medication adherence RR at 1.10 [1.00, 1.22], P=.06.

Conclusions: Digital interventions may improve healthy behavioral factors (PA, healthy diet, and medication adherence) and are even more potent when used to treat multiple behavioral outcomes (eg, medication adherence plus). However, they did not appear to reduce unhealthy behavioral factors (smoking, alcohol intake, and unhealthy diet) and clinical outcomes (BMI, triglycerides, diastolic and systolic BP, and HbA_{1c}).

(JMIR Mhealth Uhealth 2021;9(3):e21061) doi:10.2196/21061

KEYWORDS

digital technologies; mHealth; eHealth; risk factors; cardiovascular diseases; telehealth; cardiac rehabilitation; behavior; systematic review; meta-analysis; mobile phone

Introduction

Background

Cardiovascular diseases (CVDs), including coronary heart disease (CHD), stroke, and peripheral vascular disease, remain one of the most common causes of early death and disability worldwide, with 17.9 million deaths and 422.7 million cases each year [1]. In 2017 alone, there were approximately 1.7 million inpatient episodes of CVDs in the United Kingdom [2]. This imposes a heavy burden on individuals and the society, accounting for £19 billion in public expense, 7.4 million disabilities, and 167,000,000 deaths in 2019 [1].

Although there are genetic, demographic, and environmental causes of CVDs [3], approximately 50% of CVD risk is attributable to modifiable lifestyle factors such as obesity, diabetes, inactivity, and smoking [4]. However, despite widespread education and personal knowledge, many individuals fail to adequately modify these risk factors, even after a cardiovascular event with cardiac rehabilitation care center support [4]. Failure to address this challenge (ie, a change from cognitive insight to manifest action) results in patients remaining at a higher risk of future cardiovascular events with associated personal, social, and economic costs.

There could be several reasons for the challenges in the personalized management (ie, modification) of CVD and other chronic disease risk factors. These include care center accessibility and outpatient mobility and morbidity, comprehensibility, and retainability [5]. For these reasons, lifestyle risk factor management (particularly low physical activity [PA] and obesity) remains suboptimally addressed in CVD outpatients [6]. These show that although there is a modest success with rehabilitation care center interventions, the technical reach of this population-based approach is limited in its ability to bring about a significant sustainable change in exposed individuals [5].

To achieve a sustainable change, social construct strategies (eg, self-management, motivation, perceived benefits, etc) embedded in behavioral change interventions have shown health benefits in chronic disease risk factor management [7]. Their substantial contribution to changes in health behaviors suggests a worthy consideration in behavioral health interventions at the individual level. However, in the context of a population-based health behavior change in rehabilitation care centers, there are limitations in the ability of social constructs to make an individual cope sustainably with its strategies at the personal level [6].

The emergence of digital health technologies (eg, the internet, phone apps and devices, and wearable sensors for telemedicine, web browsing, emailing, text messaging, monitoring) in the health care sector [8], designed to manage and monitor chronic disease lifestyle factors, has shown potential in personalized chronic disease lifestyle factor modification [9]. This potential is based on evidence that healthy lifestyle factors are behavior-specific, measurable, and modifiable [10]. Due to the commercial drive and attributed qualities, many of these technologies and devices have been continually used in cardiac rehabilitation care centers, and even instead of care [11,12]. However, despite their popularity and potential, these technologies lack evidence summary of secondary prevention of clinically relevant outcomes, which result from behavior change in CVD outpatients, especially at a personalized level [12,13].

Objectives

The primary objective of this systematic review is to identify and measure the effectiveness of digital technology (eg, mobile phones, the internet, software applications, wearables, etc) interventions in randomized controlled trials (RCTs) and determine which behavior change constructs were effective at achieving risk factor modification among patients with CVD.

Methods

Study Design

This study is a systematic literature review and meta-analysis of RCTs designed in line with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analysis) statement standard [14]. The protocol was registered with PROSPERO protocol ID CRD42019139801 [15].

Inclusion Criteria for Considering Studies in This Review

Types of Studies

RCTs of digital interventions with a minimum of 4 weeks of intervention follow-up period were considered. Publications in English from 2000 to September 2019 were included.

Types of Participants

The considered studies focused on an adult population (≥18 years) with a minimum of 30 participants in the intervention study.

Types of Interventions

The considered intervention types were digital intervention only versus usual care or digital intervention plus usual care versus usual care. The studies were required to be based on a well-defined CVD risk factor modification measurement function and intention to modify health behavior in outpatients diagnosed or treated for CVD only or with comorbidities using a named digital device.

Types of Outcome Measures

Clinical and behavioral outcomes were measured at baseline and endpoint. All outcome hypothesis testing was set at 95% CI and a 2-tailed *P* value of .05 level of statistical significance. Behavioral outcomes included PA (physical inactivity [PI], sedentary lifestyle), food intake (diet), smoking, alcohol intake, and medication adherence. Clinical outcomes included BMI, cholesterol levels (total cholesterol [TC], high-density lipoproteins [HDLs], low-density lipoproteins [LDLs], and triglycerides [TGs]), blood pressure (diastolic BP and systolic BP), and blood sugar levels (HbA_{1c}), which are measures of obesity, hypercholesterolemia, hypertension, and diabetes, respectively.

Exclusion Criteria

Studies related to non-RCTs, nondigital interventions, and journal papers not published in English were excluded. In addition, studies whose populations shared modifiable risk factors with CVD but were not diagnosed with a named CVD; studies with nonclinical or nonbehavioral outcomes, such as studies with genetic outcomes and studies directly measuring hospital or staff service efficiency, etc; and healthy population or intensive care studies were excluded.

No analysis was conducted on data of outcomes from subgroups (endpoint to endpoint) within the population of the included studies.

Search Methods for Identification of Studies

A single 6S pyramid systematic literature search strategy was developed (Table S3 in the Multimedia Appendix 1). This was run on Ovid Medline and Ovid Embase, Web of Science (Core Collection), Scopus, Cochrane Library, and on the following databases in EBSCOHost: CINAHL, Psych Info, Health Source, Open Dissertation, Psych Article, and Business Source Elite. Filters were used to narrow searches to studies using RCT methodology and those written in the English language and from the year 2000 onward. The year limit was applied, in line with the World Health Organization's release of the document on the approach to digital health strategies [16], as the start of the mass availability and use of digital technologies, making pre-2000 literature less relevant.

Two independent reviewers (AS and RP) were involved in a thorough search strategy build-up and study extraction to identify potentially relevant publications. References and citations were also searched. Where an abstract did not provide sufficient precision to meet the selection prerequisite, the article was reserved for full-text review. Relevant articles retrieved for full-text reviews were independently evaluated (AS and MG). The consensus to include or exclude a trial was reached based on study design, method, population demography, intervention mechanism, and study outcomes.

Data Collection and Analysis

Data Collection

The PRISMA search protocol [14] was followed, with all extracted data subsequently managed using the Mendeley Desktop reference manager software (Elsevier). Publication search outcomes were imported in .ris format into the Mendeley Desktop and partitioned based on the search database source. Imported publications were autochecked for duplicates using the software, and a further manual, independent duplication check was carried out (AS and YD). Publication papers were title-read, abstract-read, and full-text read based on the inclusion and exclusion preselection criteria. Selected journal papers were read for data synthesis and analysis.

Data Extraction and Management

Data were extracted into a preset Excel (Microsoft Corporation) worksheet. The data extraction process was performed independently (AS) using predetermined variables and then validated accordingly (MG). Data extracted included population demographics (mean age, sex, size, and CVD diagnosis) description of the study (authors, year of publication, country, intervention acronym, digital device, intervention type, trial protocol registration, design, and duration), behavioral change context (change technique and risk factors), and clinical study outcomes (outcome measures, outcome units, mean baseline measurements, mean outcome measurements, *P* values, and SDs). Authors of studies with insufficient or missing outcome data were contacted for further information.

Data Analysis

All extracted data from the selected studies were analyzed (Table 1) using Review Manager 5.3 (The Cochrane Collaboration). An assessment of the risk of bias (Table 2) was carried out by 2 researchers, AS and YD, using the modified Cochrane Collaboration AUB KQ1 Risk of Bias Assessment Tool, Review Manager 5.3 (The Cochrane Collaboration) with assessment result validation by an external independent researcher. Bias quality was assessed as high, low, or unclear for individual elements from 6 items: selection, performance, attrition, reporting, proportion, outcome, and treatment efficacy. Where the attrition bias risk is high, there is more likely to be a high treatment efficacy bias except where the basis for participant dropout is a medical reason, relocation, or death. Quality assessment items were evaluated by an external assessor to validate the initial scales judged by the author. Controversial evaluation differences were discussed, and consensus was reached before the final documentation. The risk of bias across studies was assessed for each analyzed outcome for publication bias reporting. Results were generated with meta-analysis data for each outcome and are presented in the Results section.

The review authors considered the variations in outcome measurement across studies by applying appropriate statistical methods (fixed effect and random effect) using the Inverse-Variance and Mantel-Haenszel (DerSimonian and Laird) models to generate meta-analytic estimates of treatment effect

using the Review Manager 5.3 software. Differences in effects were examined by comparing digital with usual care. The weighted mean difference (MD) or standardized mean difference (SMD) was calculated for continuous data using the inverse variance statistical method. Relative risks were calculated for dichotomous data using the Mantel-Haenszel statistical method (The Cochrane Collaboration). Provision for variations among the included studies was made using the random effect meta-analysis model in analyzing all included studies. The heterogeneity statistic I^2 was calculated to describe the percentage of variation among the studies. Hypothesis testing was set at a 2-tailed 0.05 level of significance and a 95% CI. No analysis was conducted on the data of outcomes from subgroups within the population of the included studies.

Sensitivity analyses were proportionately conducted on outcomes to check the cumulative effects of the publication year, participant size, efficacy, and category of intervention (risk factors and digital intervention) on statistical significance. For food intake, studies with interventions targeting healthy diet and studies targeting unhealthy diet were analyzed separately to provide a clearer insight into treatment effects. Studies with treatment for medication adherence were analyzed separately for (1) other risk factor treatments plus medication adherence using the SMS text messaging intervention only, (2) medication adherence treatment only (with no other risk factor) using the SMS text messaging intervention alone, (3) other risk factor treatment plus medication adherence treatment using non-SMS text messaging intervention, and (4) medication adherence treatment with SMS text messaging intervention only. The results are presented and discussed in the Results and Discussion sections, respectively.

Table 1. Summary of meta-analysis results[a].

Outcomes or subgroups	Number of studies (N=25), n (%)	Participants	Statistical methods	Effect estimates (95% CI)	*P* value
BMI	10 (40)	2558	MD[b] (IV, random, 95% CI)	−0.37 (−1.20 to 0.46)	.38
Total cholesterol	9 (36)	1783	SMD[c] (IV, random, 95% CI)	−0.29 (−0.44 to −0.15)	<.001
High-density lipoprotein	9 (36)	1783	SMD (IV, random, 95% CI)	−0.09 (−0.19 to 0.00)	.05
Low-density lipoprotein	12 (48)	3431	SMD (IV, random, 95% CI)	−0.18 (−0.33 to −0.04)	.01
Triglycerides	8 (32)	1660	SMD (IV, random, 95% CI)	−0.07 (−0.24 to 0.11)	.28
Diastolic BP[d]	11 (44)	2460	SMD (IV, random, 95% CI)	−0.06 (−0.20 to 0.08)	.43
Systolic BP	12 (48)	3283	SMD (IV, random, 95% CI)	−0.03 (−0.18 to 0.13)	.74
Physical activity	14 (56)	3015	SMD (IV, random, 95% CI)	0.23 (0.11 to 0.36)	<.001
Alcohol consumption	4 (16)	651	SMD (IV, random, 95% CI)	−0.16 (−1.43 to 1.10)	.80
Blood sugar, HbA_{1c}[e]	2 (8)	380	RR[f] (M-H, random, 95% CI)	1.04 (0.40 to 2.70)	.94
Physical inactivity	4 (16)	1054	RR (M-H, random, 95% CI)	0.54 (0.39 to 0.75)	<.001
Food intake (diet)	6 (24)	716	RR (M-H, random, 95% CI)	0.79 (0.66 to 0.94)	.007
Health diet	3 (12)	173	RR (M-H, random, 95% CI)	0.70 (0.55 to 0.89)	.004
Unhealthy diet	3 (12)	185	RR (M-H, random, 95% CI)	0.90 (0.68 to 1.19)	.47
Smoking	11 (44)	2916	RR (M-H, random, 95% CI)	0.87 (0.67 to 1.13)	.30
Medication adherence	11 (44)	2710	RR (M-H, random, 95% CI)	1.10 (1.00 to 1.22)	.06
Medication adherence (multiple treatment)	5 (20)	758	RR (M-H, random, 95% CI)	1.07 (1.01 to 1.14)	.02

[a]Summary of analyzed data.

[b]MD: mean difference.

[c]SMD: standard mean difference.

[d]BP: blood pressure.

[e]HbA_{1c}: Hemoglobin A_{1c} blood sugar.

[f]RR: risk ratio.

Table 2. Risk of bias in included studies[a].

Study (reference)	Selection bias		Performance bias	Detection bias	Attrition bias	Proportion bias	Outcome bias	Reportin bias	Treatmen efficacy
	Random sequence generation	Allocation concealment	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Groups balanced at baseline	Groups received same intervention	Selective reporting	Intention-to-treat analysis
Akhu-Z et al, 2016 [17]	Low	Unclear	Unclear	Low	High	Low	Low	Low	High
Chow et al, 2015 [18]	Low	Low	Low	Low	High	Low	High	Low	High
Dale et al, 2015 [19]	Low	Low	Low	High	Low	High	High	Low	Low
Devi et al, 2014 [20]	Low	Low	Low	High	High	Low	Low	High	Unclear
Frederix et al, 2015 [21]	Low	Low	Low	Low	High	Low	Low	Low	Unclear
Hawkes et al, 2012 [22]	Low	High	Low	Low	High	Low	Low	High	Unclear
Johnston et al, 2016 [23]	Low	Unclear	Unclear	Unclear	High	High	High	High	Unclear
Kamal et al, 2015 [24]	Low	Low	Low	Low	High	Low	High	Low	High
Khonsari et al, 2015 [25]	Low	Unclear	High	Unclear	Low	Low	High	Low	Low
Kraal et al, 2014 [26]	Low	Low	Unclear	Unclear	Low	Low	High	Low	Low
Lear et al, 2014 [27]	Low	Low	Low	Low	High	Low	Low	Low	Low
Maddison et al, 2014 [28]	Low	Low	Low	Low	High	Low	High	Low	Unclear
Ogren et al, 2018 [29]	Low	Low	High	Unclear	Low	Low	Low	Unclear	High
Pandey et al, 2014 [30]	Unclear	Unclear	Low	Low	Unclear	Low	Low	Unclear	Unclear
Park et al, 2013 [31]	Low	Low	High	Low	Low	Unclear	Low	Low	Low
Quilici et al, 2012 [32]	Low	Low	Unclear	Unclear	High	Low	Low	Low	Unclear
Redfern et al, 2009 [33]	Low	Low	Low	Low	Low	Low	Low	Low	Low
Reid et al, 2011 [34]	Low	Low	Low	Low	High	Low	Low	Low	Low
Southard et al, 2003 [35]	Low	Low	Low	Low	Low	Low	Low	Low	Low
Tiede et al, 2017 [36]	Low	Low	High	Unclear	High	High	Low	Low	High
Vale et al, 2002 [37]	Low	High	High	Low	High	Low	Low	High	Low
Vemooij et al, 2012 [38]	Low	Low	Unclear	Unclear	Low	Low	High	Low	Low
Wan et al, 2016 [39]	Low	Low	Low	Low	High	Low	Low	Low	Low
Widmer et al, 2017 [40]	Low	Low	Low	Low	Low	High	High	Low	Unclear
Zheng et al, 2019 [41]	Low	High	Low	Low	Low	Low	High	Low	Low

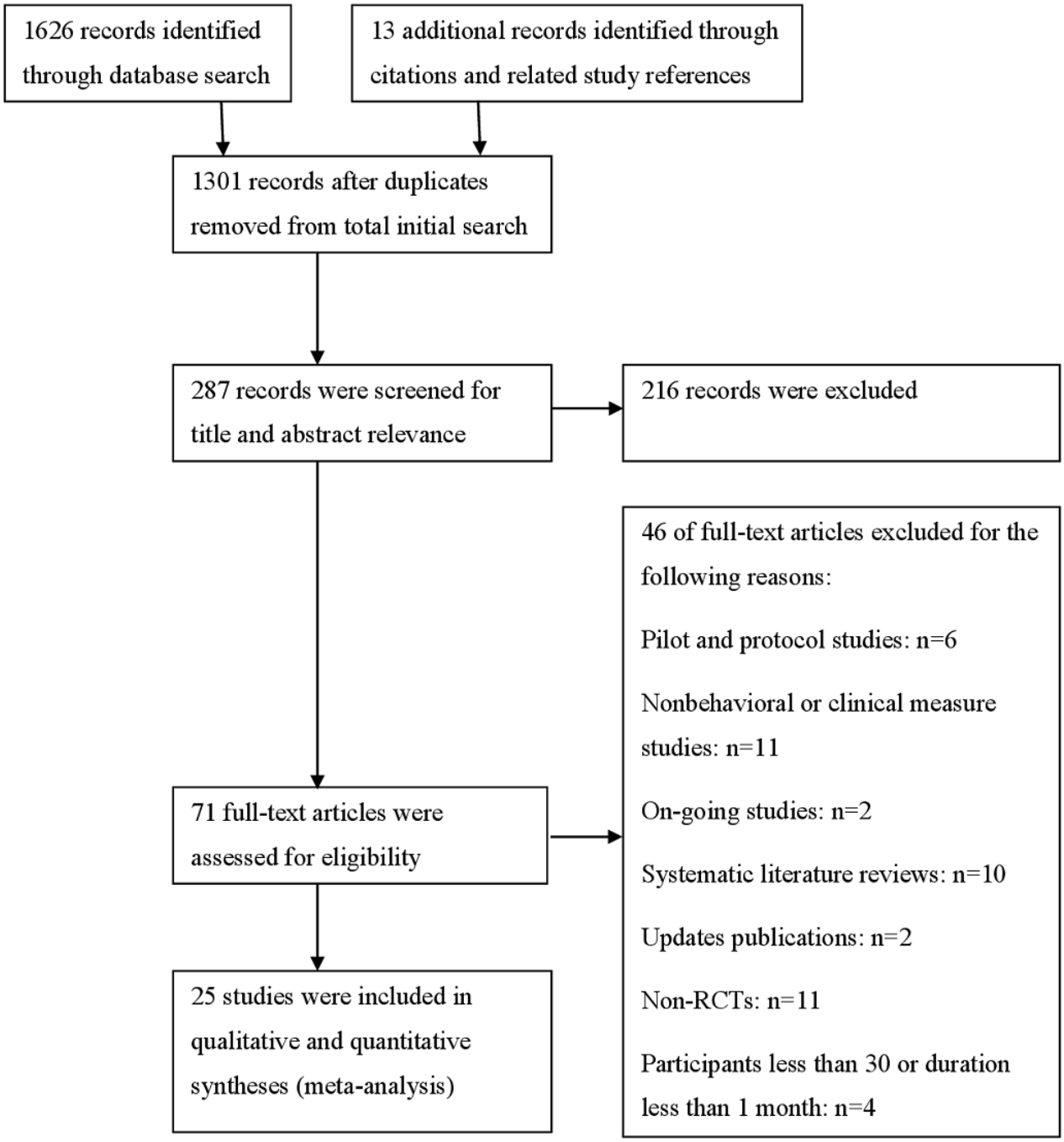
[a]Risk of bias: review authors' judgments about each risk of bias item for the included studies.

Results

Search Results

The search retrieved 1626 papers with the auto-removal of 326 duplicates. A total of 35 papers remained after applying inclusion and exclusion criteria. Ten papers were excluded because they were systematic literature reviews but not RCTs. A final count of 25 papers was considered for the review: 12 from the database searches and 13 from references, citations, and gray literature (Figure 1). The included studies are listed in the table of included studies (Table S4 in Multimedia Appendix 1), and excluded studies are listed in the table of excluded studies (Table S5 in Multimedia Appendix 1) with reasons.

Figure 1. Database search flowchart. RCT: randomized controlled trial.



Study Characteristics

Studies are described by their common characteristics, which include population demography, digital technologies and brands, intervention mechanisms and behavioral change constructs, types of CVD, and general characteristics. Table S6 (Multimedia Appendix 2) provides a detailed summary of the reviewed studies.

Population Demography

The included studies had a total participant count of 5,779 at baseline, with a mean age of 60.03 years (SD 2.73) and a male proportion of 75.22% (4347/5779). A geographical analysis of the included studies identified evenly distributed locations of studies on a global scale, with countries spanning Europe (5 studies), Middle East (2 studies), Asia (3 studies), Northern America (6 studies), Scandinavia (2 studies), Australasia (6 studies), and the United Kingdom (1 study).

Digital Technologies and Brands

Cardiovascular digital interventions were delivered using devices such as cell phones, smartphones, personal computers (laptops and desktops), and wearables. Technologies included the internet, software applications, and mobile sensors. Intervention device brand names such as Personal Health Assistant, PHA, FIT@Home, HeartLinks, SUPPORT,

SMS4Stroke, ProActive Heart, Text4Heart, CHAT, CardioFit, HEART, ActivateYourHeart, MEMS, vCRP, COACH, CHOICE, and TBHC were recorded.

Intervention Mechanisms and Behavioral Change Constructs

Intervention mechanisms (ie, the digital strategy plus behavioral construct) were based on online support, telerehabilitation, telemonitoring, and online coaching. The interventions included major behavioral change constructs such as cognition, follow-up, goal setting, record keeping, perceived benefit, persuasion, social engagement (virtual), personalization (or customization), rewards and incentives, support, and self-management.

CVD types: Diagnosed CVDs included CHD, coronary artery disease, myocardial infarction, acute coronary syndrome, angina, atherosclerosis, heart failure, transient ischemic attack, and stroke. Four studies [17,27,35,42] were not specific about CVD diagnosis in the study population.

General Characteristics

Study follow-up ranged from 1 to 4 (9 studies), 6 (12 studies), 12 (3 studies), and 24 months (1 study), with 6 months being the most frequent duration of the follow-up period for interventions. No data on outcomes from subgroups within the populations of the included studies were considered in the analysis.

The main units of outcome measurements were kg/m^2 (BMI), mg/dL, and mmol/L (TC, HDL, low-density lipoprotein [LDL], and TGs, mmHg (diastolic blood pressure [DBP] and systolic blood pressure [SBP]), min/week (PA and PI), percentage, % (Hemoglobin A_{1C} blood sugar—HbA_{1c}, alcohol, smoking, and food intake), and Morisky Medication Adherence Scale for Medication Adherence 8. In the treatment context, all the intervention studies had been administered either as digital intervention versus usual care (15 studies) or digital intervention and usual care versus usual care (10 studies).

Synthesis of Results

The use of digital intervention compared with usual care significantly modified all CVD risk factors except BMI, TG, DBP, SBP, HbA_{1c}, alcohol intake, smoking, and medication adherence. A detailed summary of these findings is presented in Table 1.

Summary of Results

Effect estimates (MD, SMD, and RR) were significant and in favor of digital interventions for TC, HDL, LDL, PA, PI, and food intake.

Clinical Outcomes

The BMI outcome (Figure 2) reported MD was estimated at −0.37 (95% CI –1.20 to 0.46, *P*=.38). The TC outcome (Figure 3) reported an SMD estimated at −0.29 (95% CI –0.44 to –0.15, *P*<.001). The HDL outcome (Figure 4) reported an SMD estimated at −0.09 (95% CI −0.19 to 0.00, *P*=.05). The LDL outcome (Figure 5) reported an SMD estimated at –0.18 (95% CI –0.33 to –0.04, *P*=.01). The TG outcome (Figure 6) reported an SMD estimated at −0.10 (95% CI –0.28 to 0.08), *P*=.28. Diastolic and systolic BP outcomes (Figures 7 and 8) reported SMDs estimated at –0.06 (95% CI −0.20 to 0.08, *P*=.43) and −0.03 (95% CI −0.18 to 0.13, *P*=.74), respectively. The HbA_{1c} outcome (Figure 9) reported an RR estimated at 1.04 (95% CI 0.40 to 2.70, *P*=.94). A summary of the clinical outcome findings is presented in Figures 2-9.

Figure 2. Outcomes of the examined studies for BMI.

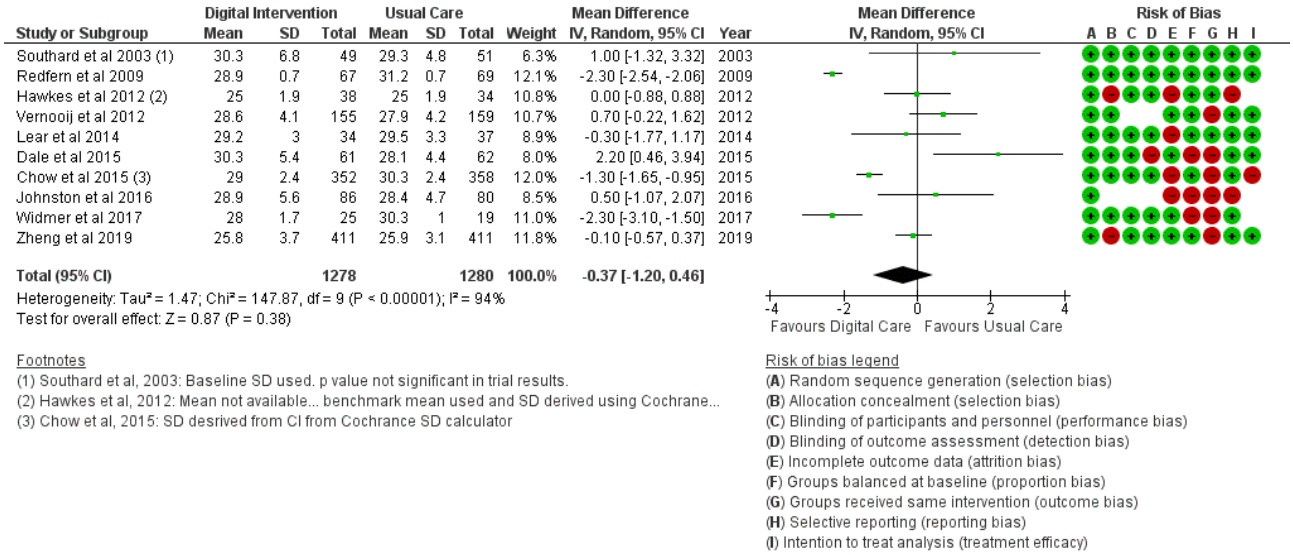
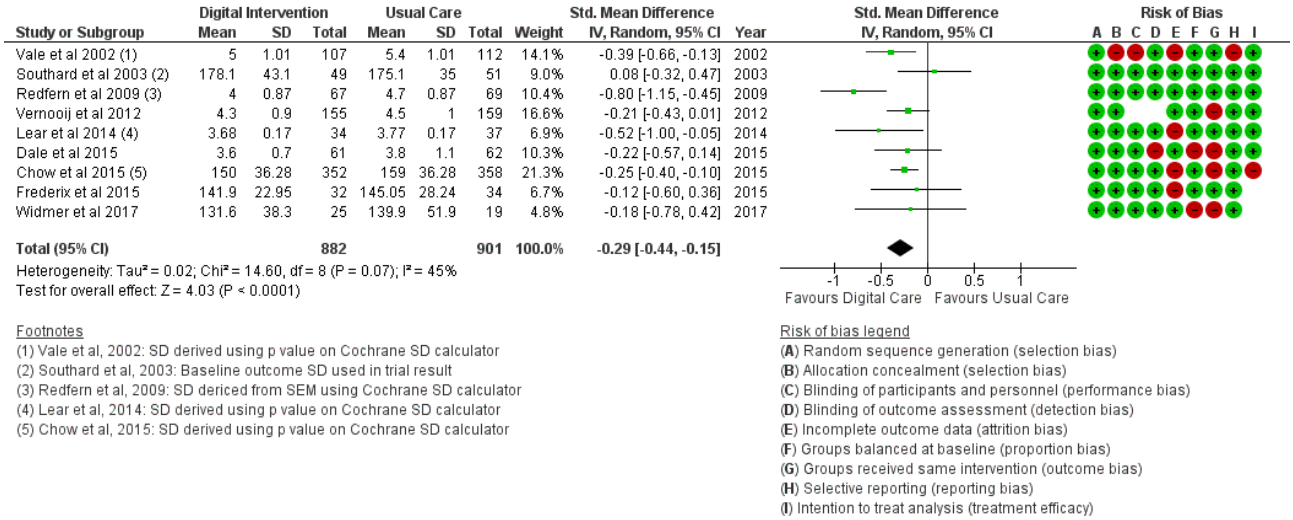


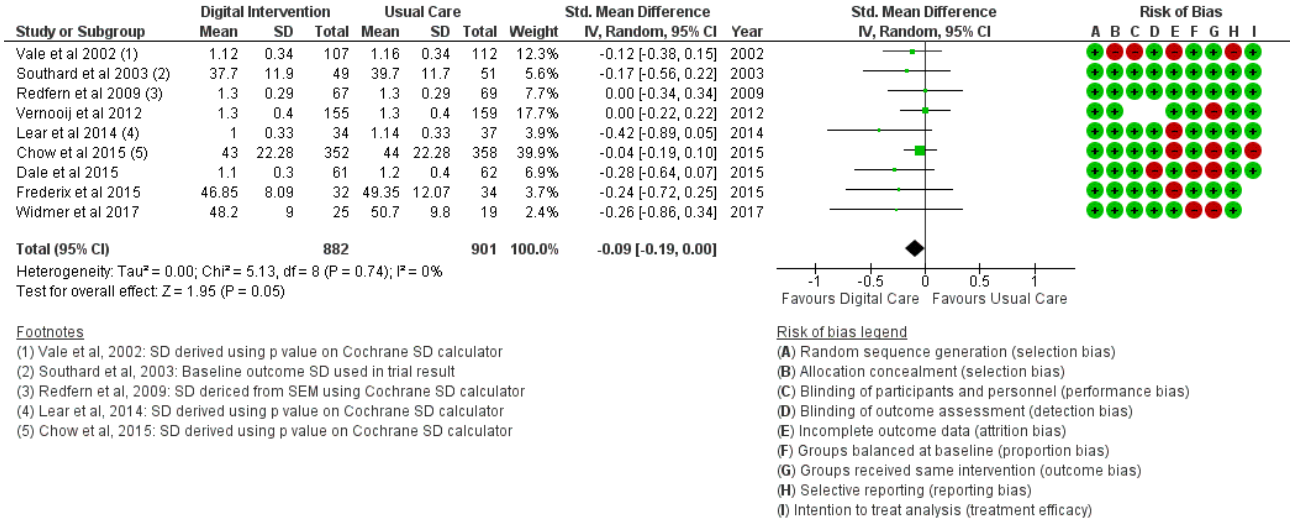
Figure 3. Outcomes of the examined studies for total cholesterol.



Footnotes
(1) Vale et al, 2002: SD derived using p value on Cochrane SD calculator
(2) Southard et al, 2003: Baseline outcome SD used in trial result
(3) Redfern et al, 2009: SD dericed from SEM using Cochrane SD calculator
(4) Lear et al, 2014: SD derived using p value on Cochrane SD calculator
(5) Chow et al, 2015: SD derived using p value on Cochrane SD calculator

Risk of bias legend
(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Blinding of participants and personnel (performance bias)
(**D**) Blinding of outcome assessment (detection bias)
(**E**) Incomplete outcome data (attrition bias)
(**F**) Groups balanced at baseline (proportion bias)
(**G**) Groups received same intervention (outcome bias)
(**H**) Selective reporting (reporting bias)
(**I**) Intention to treat analysis (treatment efficacy)

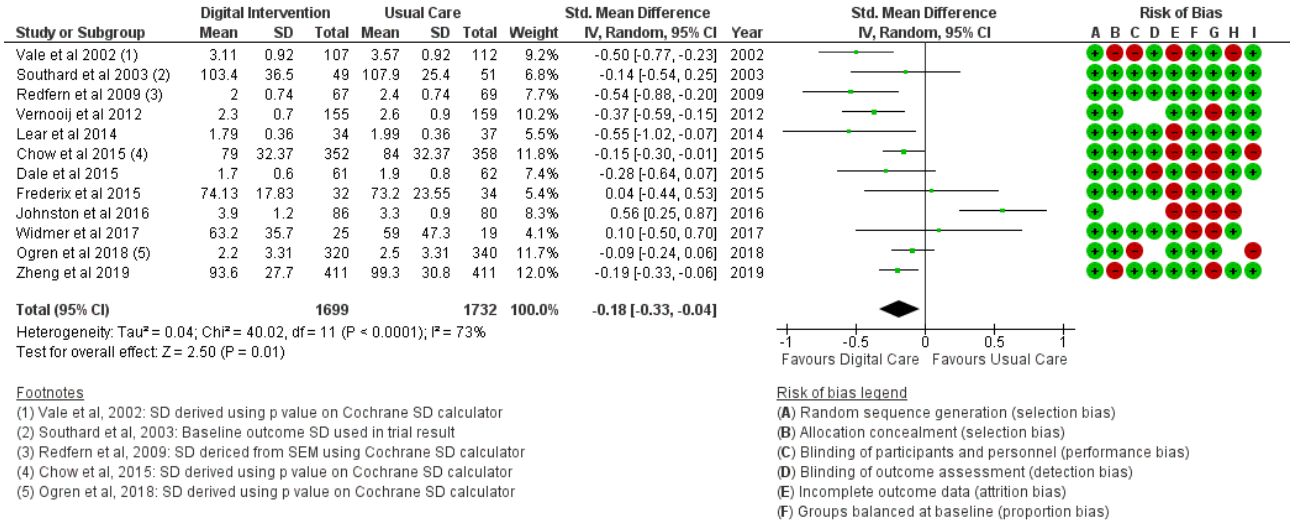
Figure 4. Outcomes of the examined studies for high-density lipoprotein.



Footnotes
(1) Vale et al, 2002: SD derived using p value on Cochrane SD calculator
(2) Southard et al, 2003: Baseline outcome SD used in trial result
(3) Redfern et al, 2009: SD dericed from SEM using Cochrane SD calculator
(4) Lear et al, 2014: SD derived using p value on Cochrane SD calculator
(5) Chow et al, 2015: SD derived using p value on Cochrane SD calculator

Risk of bias legend
(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Blinding of participants and personnel (performance bias)
(**D**) Blinding of outcome assessment (detection bias)
(**E**) Incomplete outcome data (attrition bias)
(**F**) Groups balanced at baseline (proportion bias)
(**G**) Groups received same intervention (outcome bias)
(**H**) Selective reporting (reporting bias)
(**I**) Intention to treat analysis (treatment efficacy)

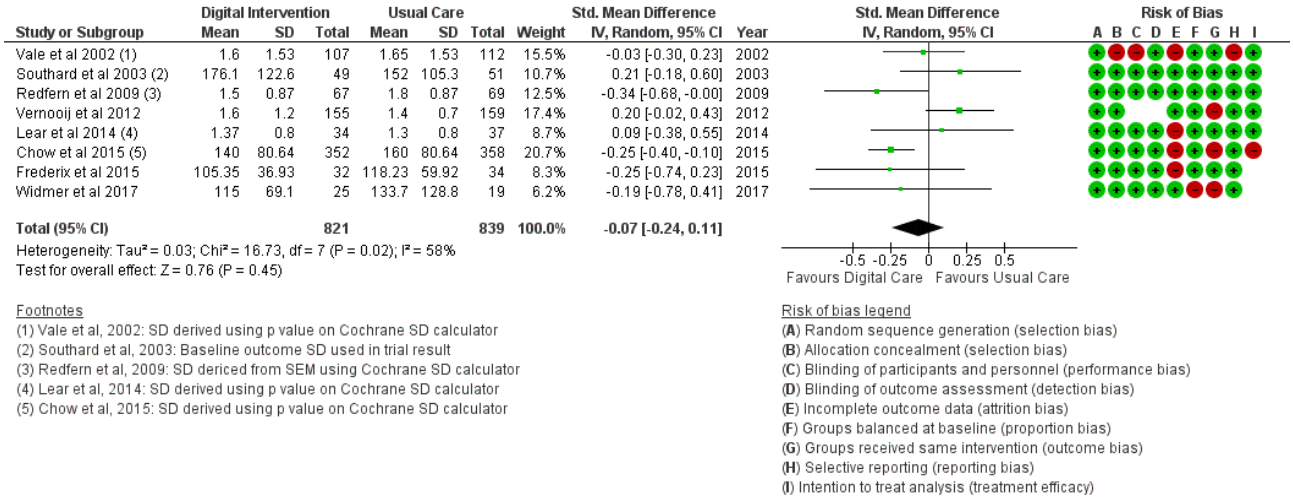
Figure 5. Outcomes of the examined studies for low-density lipoprotein.



Footnotes
(1) Vale et al, 2002: SD derived using p value on Cochrane SD calculator
(2) Southard et al, 2003: Baseline outcome SD used in trial result
(3) Redfern et al, 2009: SD dericed from SEM using Cochrane SD calculator
(4) Chow et al, 2015: SD derived using p value on Cochrane SD calculator
(5) Ogren et al, 2018: SD derived using p value on Cochrane SD calculator

Risk of bias legend
(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Blinding of participants and personnel (performance bias)
(**D**) Blinding of outcome assessment (detection bias)
(**E**) Incomplete outcome data (attrition bias)
(**F**) Groups balanced at baseline (proportion bias)
(**G**) Groups received same intervention (outcome bias)
(**H**) Selective reporting (reporting bias)
(**I**) Intention to treat analysis (treatment efficacy)

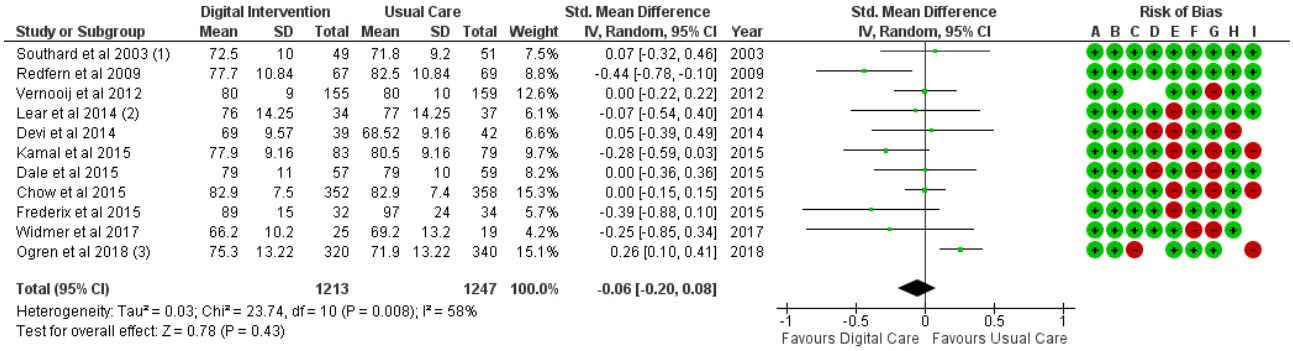
Figure 6. Outcomes of the examined studies for triglycerides.



Footnotes
(1) Vale et al, 2002: SD derived using p value on Cochrane SD calculator
(2) Southard et al, 2003: Baseline outcome SD used in trial result
(3) Redfern et al, 2009: SD dericed from SEM using Cochrane SD calculator
(4) Lear et al, 2014: SD derived using p value on Cochrane SD calculator
(5) Chow et al, 2015: SD derived using p value on Cochrane SD calculator

Risk of bias legend
(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Blinding of participants and personnel (performance bias)
(**D**) Blinding of outcome assessment (detection bias)
(**E**) Incomplete outcome data (attrition bias)
(**F**) Groups balanced at baseline (proportion bias)
(**G**) Groups received same intervention (outcome bias)
(**H**) Selective reporting (reporting bias)
(**I**) Intention to treat analysis (treatment efficacy)

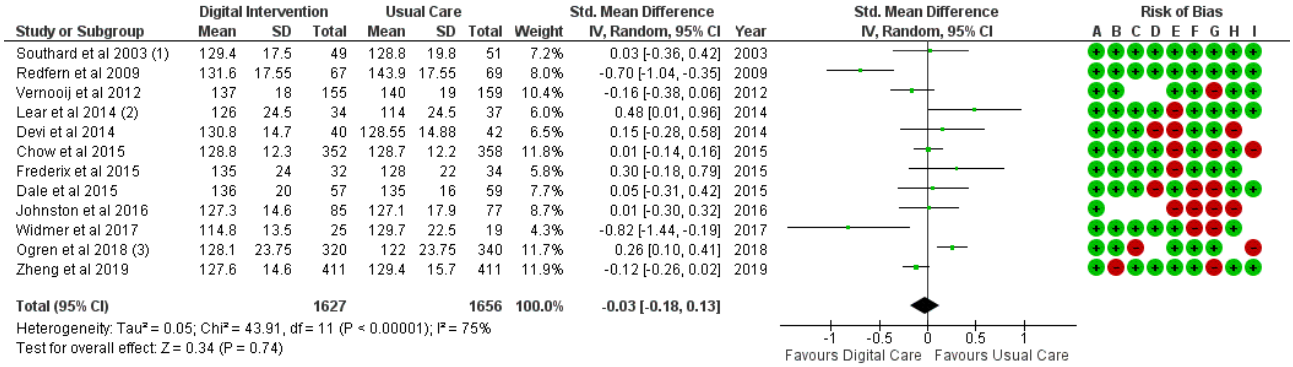
Figure 7. Outcomes of the examined studies for diastolic blood pressure.



Footnotes
(1) Southard et al, 2003: Baseline outcome SD used in trial result.
(2) Lear et al, 2014: SD derived using p value on Cochrane SD calculator
(3) Ogren et al, 2018: SD derived using p value on Cochrane SD calculator

Risk of bias legend
(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Blinding of participants and personnel (performance bias)
(**D**) Blinding of outcome assessment (detection bias)
(**E**) Incomplete outcome data (attrition bias)
(**F**) Groups balanced at baseline (proportion bias)
(**G**) Groups received same intervention (outcome bias)
(**H**) Selective reporting (reporting bias)
(**I**) Intention to treat analysis (treatment efficacy)

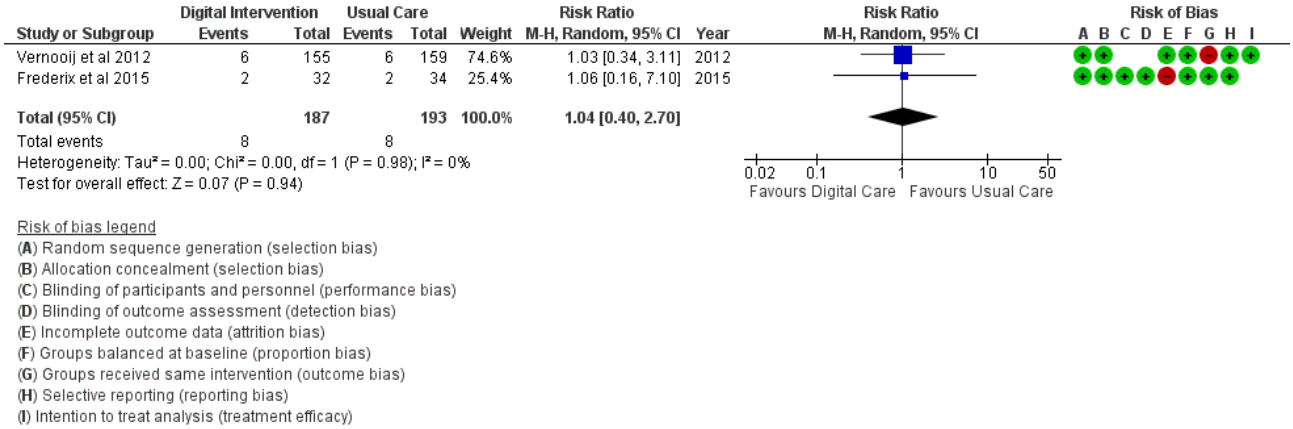
Figure 8. Outcomes of the examined studies for systolic blood pressure.



Footnotes
(1) Southard et al, 2003: Baseline outcome SD used in trial result.
(2) Lear et al, 2014: SD derived using p value on Cochrane SD calculator
(3) Ogren et al, 2018: SD derived using p value on Cochrane SD calculator

Risk of bias legend
(**A**) Random sequence generation (selection bias)
(**B**) Allocation concealment (selection bias)
(**C**) Blinding of participants and personnel (performance bias)
(**D**) Blinding of outcome assessment (detection bias)
(**E**) Incomplete outcome data (attrition bias)
(**F**) Groups balanced at baseline (proportion bias)
(**G**) Groups received same intervention (outcome bias)
(**H**) Selective reporting (reporting bias)
(**I**) Intention to treat analysis (treatment efficacy)

Figure 9. Outcomes of the examined studies for blood sugar HbA_{1c}.



Behavioral Outcomes

The PA outcome (Figure 10) reported an SMD estimated at 0.23 (95% CI 0.11 to 0.36, *P*<.001). PI (sedentary) in Figure 11 reported an RR estimated at 0.54 (95% CI 0.39 to 0.75, *P*<.001). Diet (food intake) in Figure 12 reported an RR estimated at 0.79 (0.66 to 0.94, *P*=.007). Further analysis was conducted as healthy diet targeted (Figure 13) treatment and unhealthy diet targeted (Figure 14) treatment; this is reported in the *AdditionalAnalysis* section. The alcohol intake outcome (Figure 15) reported an SMD estimated at −0.16 (95% CI −1.43 to 1.10, *P*=.80). Smoking and medication adherence outcomes (Figures 16 and 17) reported RR estimated at 0.87 (95% CI 0.67 to 1.13, *P*=.30), and 1.10 (95% CI 1.00 to 1.22, *P*=.06), respectively. A summary of the behavioral outcome findings is presented in Figures 10-21.

Figure 10. Outcomes of the examined studies for physical activity.

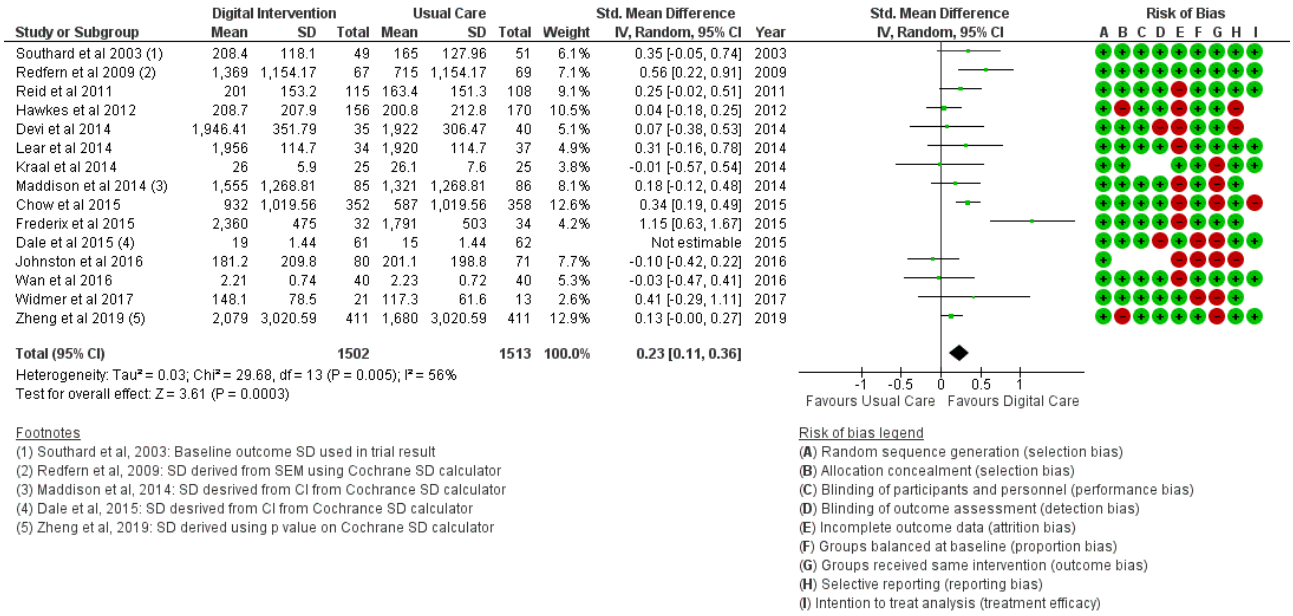


Figure 11. Outcomes of the examined studies for physical inactivity.

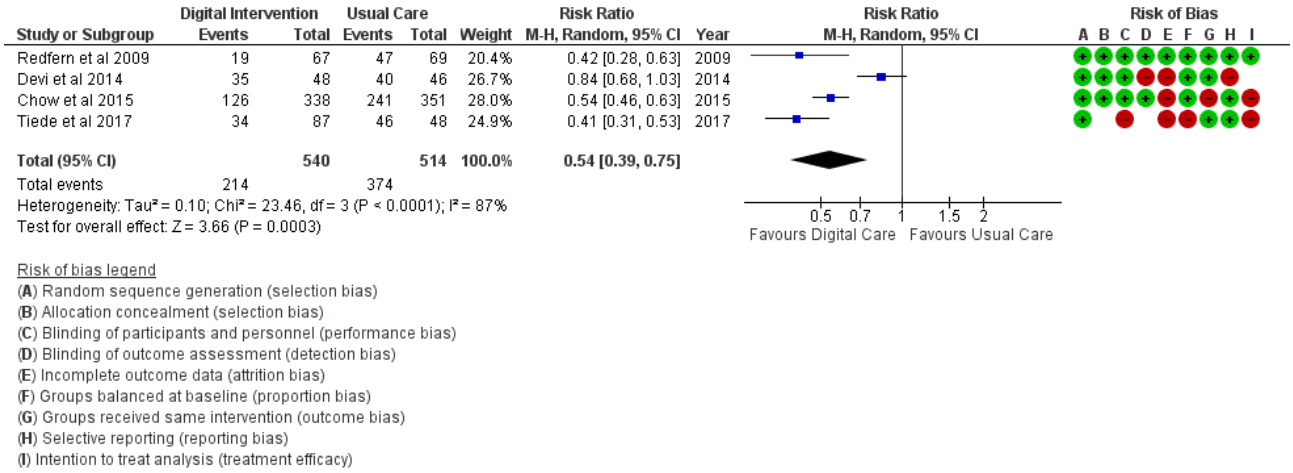


Figure 12. Outcomes of the examined studies for food intake.

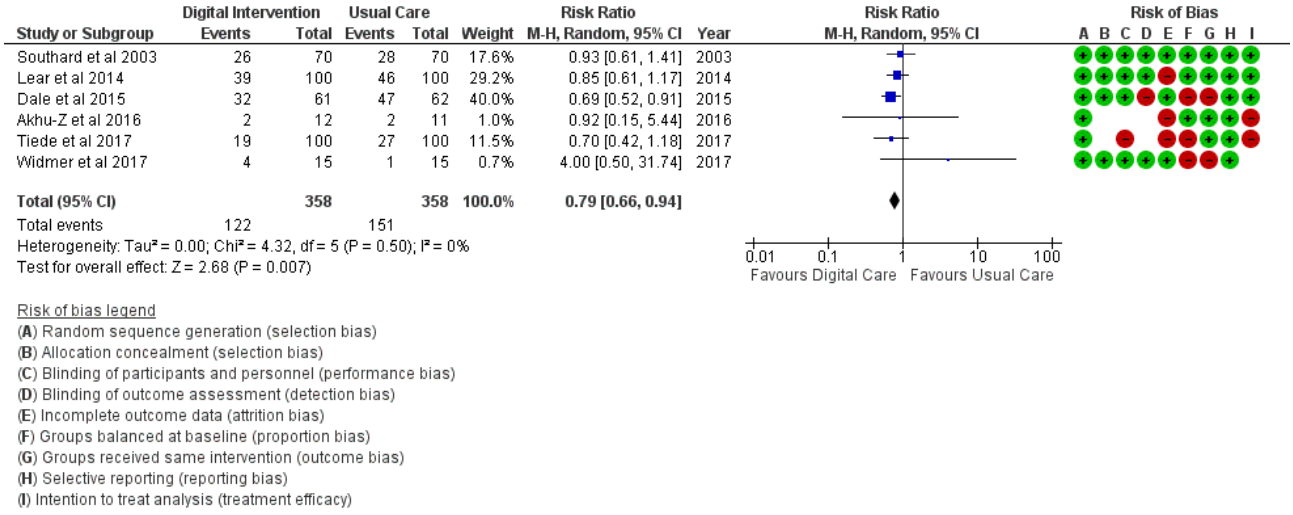


Figure 13. Outcomes of the examined studies for healthy diet.

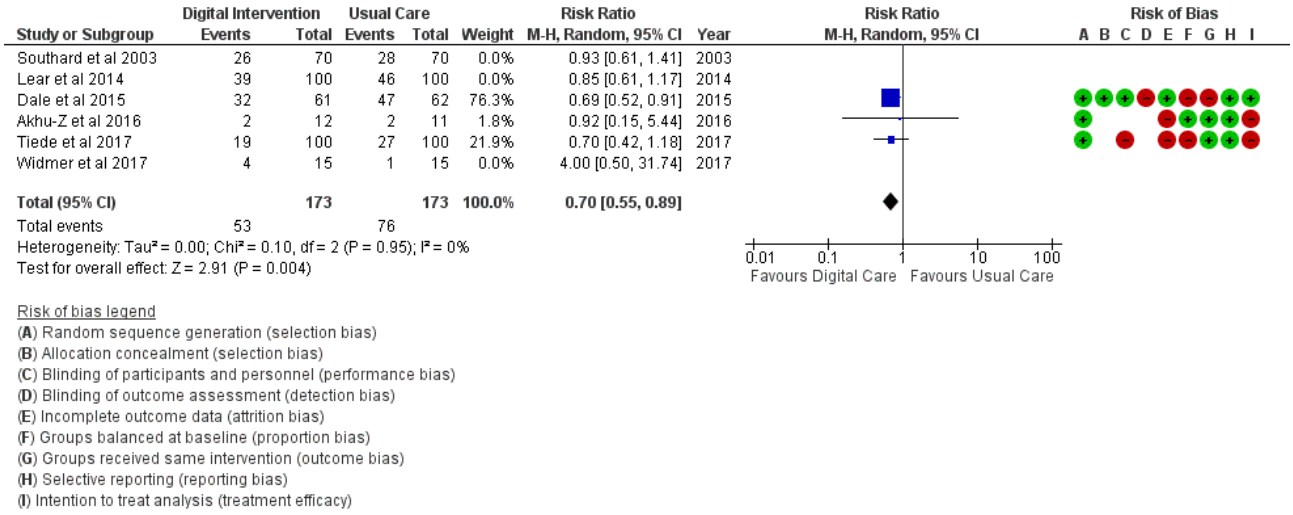


Figure 14. Outcomes of the examined studies for unhealthy food intake.

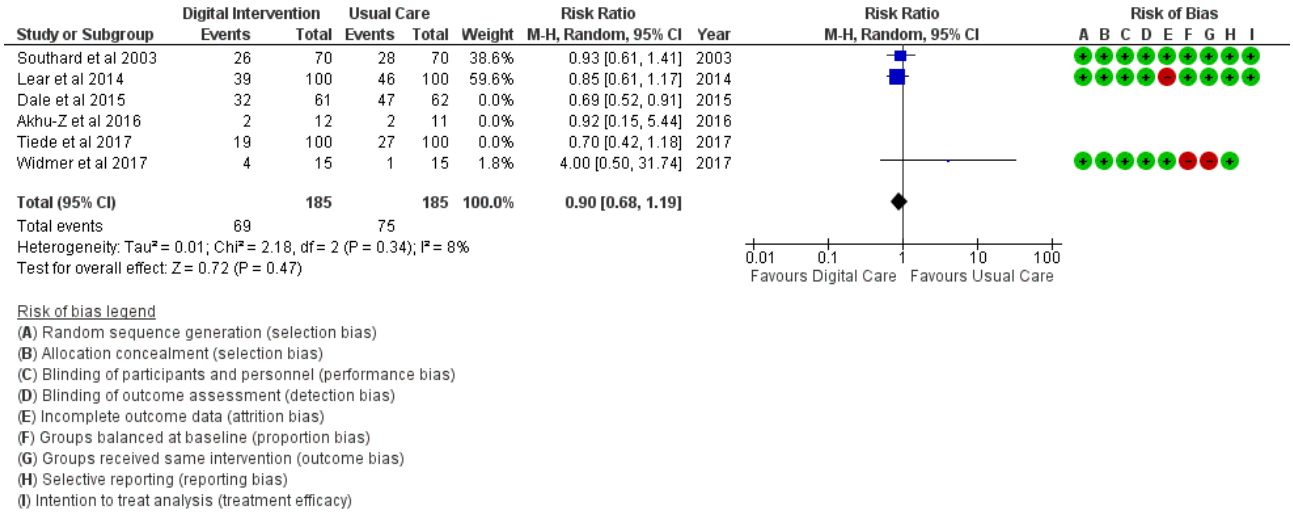


Figure 15. Outcomes of the examined studies for alcohol consumption.

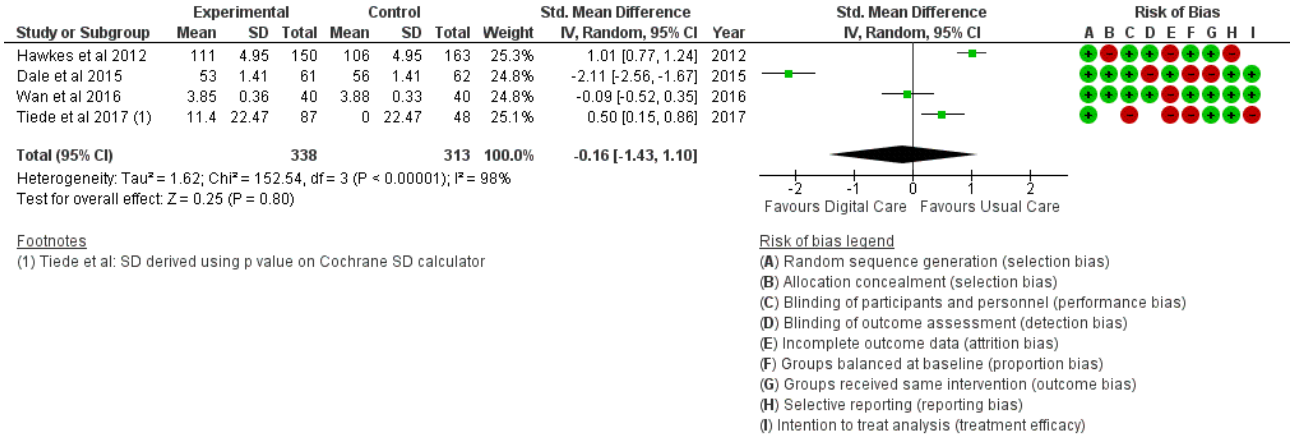


Figure 16. Outcomes of the examined studies for smoking.

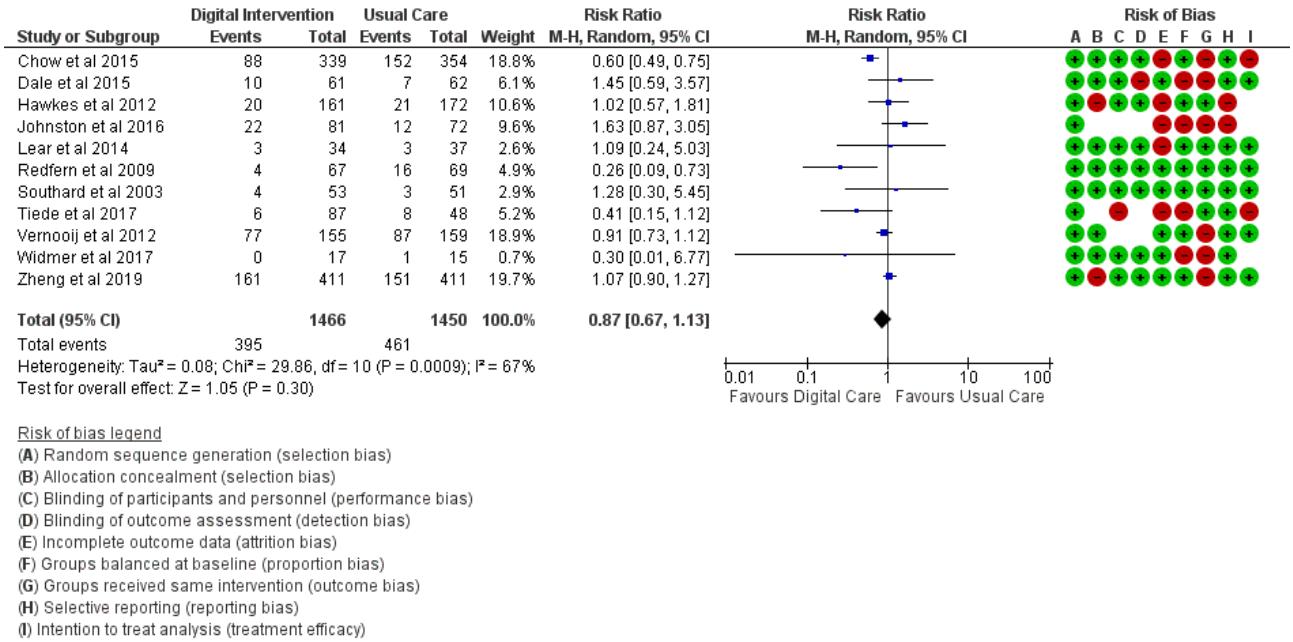


Figure 17. Medication adherence for all trials.

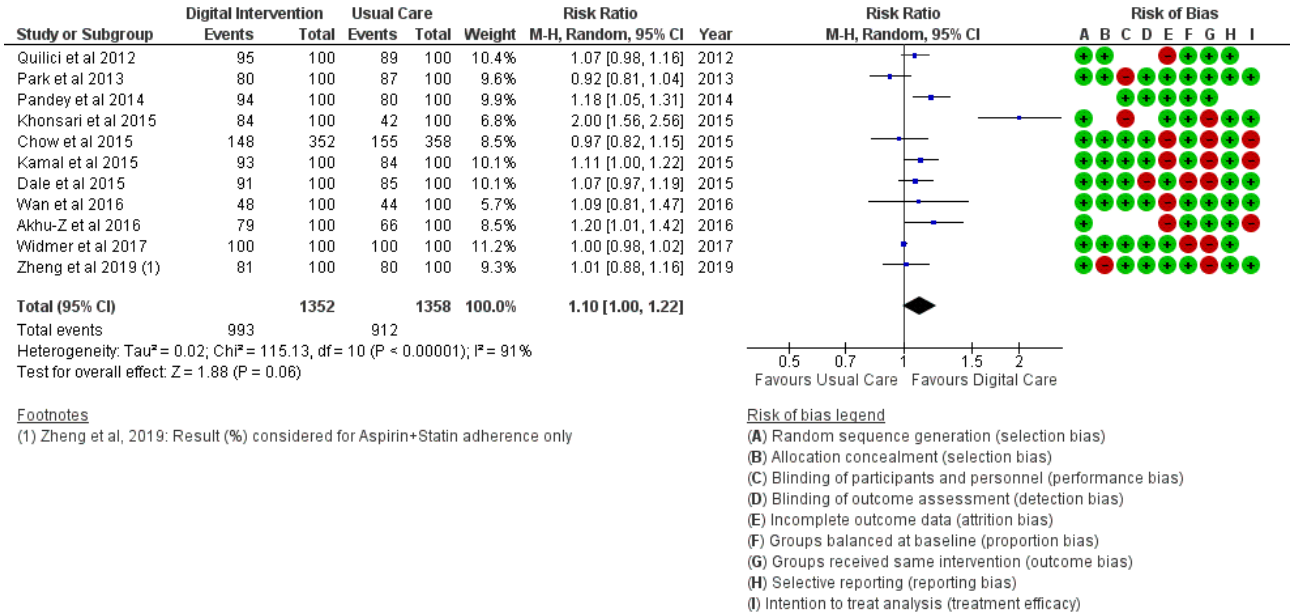


Figure 18. Medication adherence for multiple treatment with SMS.

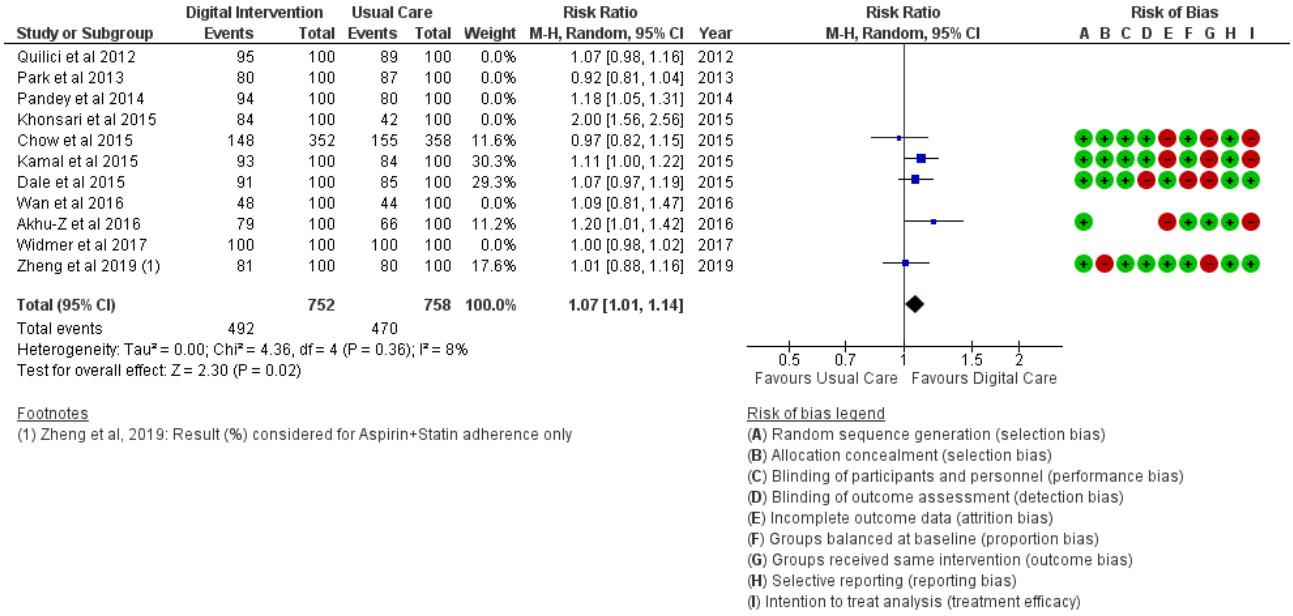


Figure 19. Medication adherence for target treatment only with SMS text message intervention.

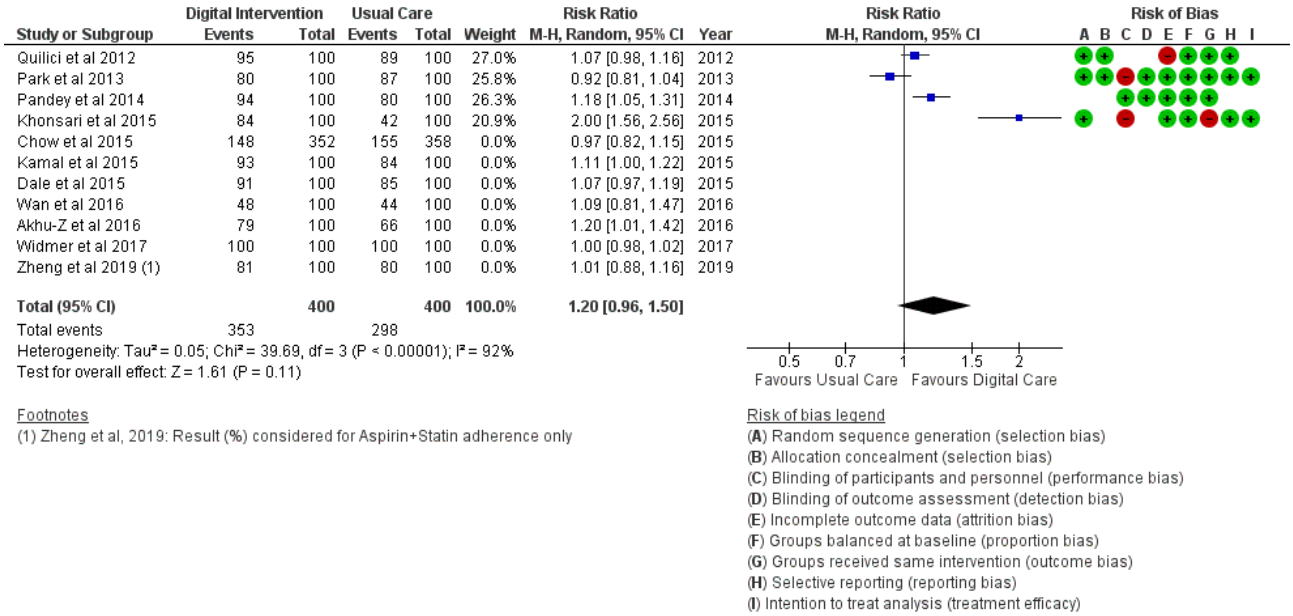


Figure 20. Medication adherence for treatment with non-sms intervention.

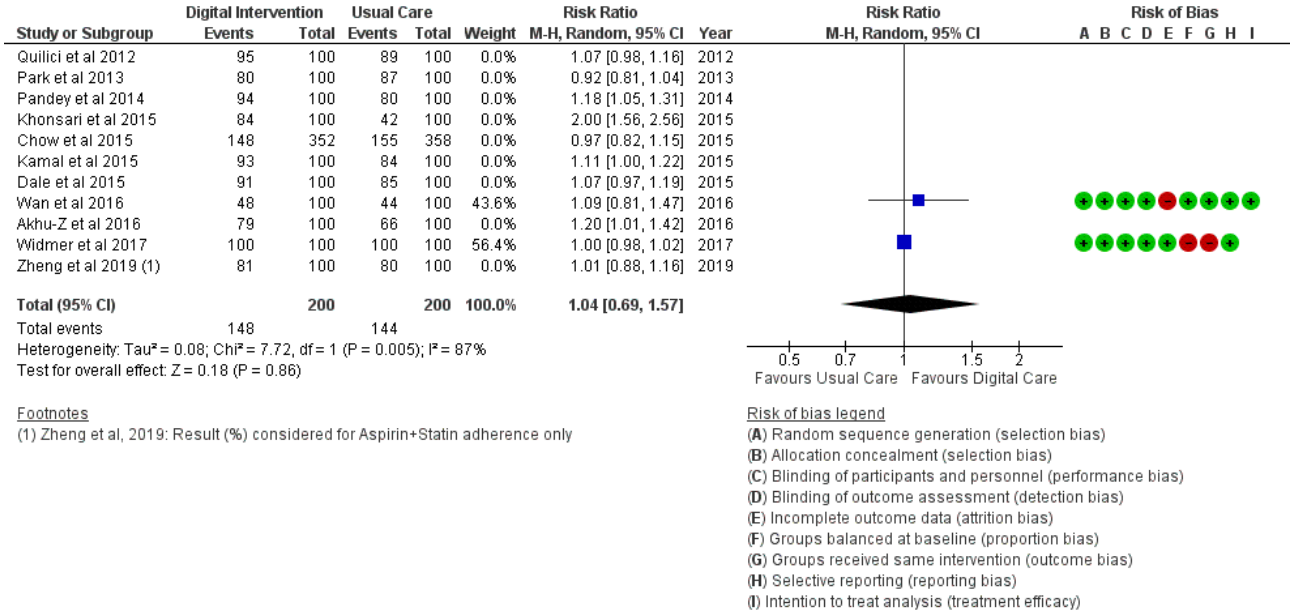
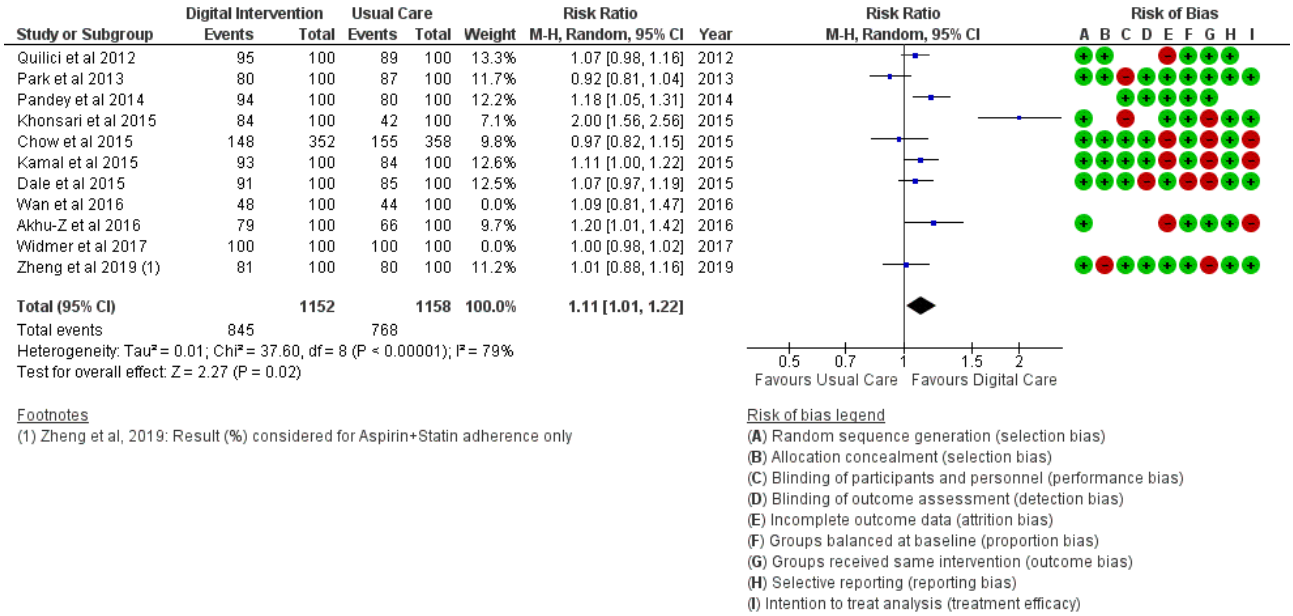


Figure 21. Medication adherence treatment for all sms intervention.



Risk of Bias in Included Studies

Table 2 provides a qualitative detail of the risk of bias assessment of the studies in the review. Proportion bias at baseline was reported in 16% (4/25) of the included studies as high risk. Intervention dropout was recorded in 32% (8/25) of the included studies at less than 10% of participants per study. Dropouts greater than 10% of study participants were recorded as high risk for treatment efficacy.

Risk of Bias Across Studies

The results of the risk of bias across studies for each outcome are presented along with the meta-analysis (Figures 2-21). Outcomes are sparsely identified with low and unclear risks for the identified risk items.

Additional Analysis

Results of the sensitivity analysis were included for the considered outcomes in Figures 13 and 14 and Figures 18-21. Sensitivity analyses proportionately conducted on outcomes to check the cumulative effects of the study population and intervention characteristics (eg, publication year, participant size, efficacy, and categories of treatment) on subgroups showed significant effects of interest in 2 outcomes as follows:

Food intake: Healthy diet targeted (Figure 13) treatment (P=.004) and unhealthy diet targeted (Figure 14) treatment (P=.47); medication adherence: medication adherence plus other risk factors (Figure 18) treatment (P=.02) and medication adherence treatment alone (P=.11), as shown in Figure 19.

Discussion

Principal Findings

This study of digital technology interventions addressing clinical and behavioral risk factor modification in people with CVDs demonstrates that not all CVD lifestyle risk factor modifications are favored by the use of digital interventions. Digital technology intervention in cardiac patients was associated with improvements in TC, HDL, LDL, PA, PI, healthy diet, and medication adherence (all P≤.05). However, there were no differences in intervention effects for BMI, TGs, BP (diastolic and systolic), blood sugar, alcohol intake, and smoking (all P>.05).

Behavioral Change Constructs and Digital Intervention Strategies

The mechanism of risk factor modification in the included studies is based on behavioral change constructs [43,44], which include self-management, feedback mechanisms, progress recording and tracking (monitoring), one-on-one or social support, persuasion, personalization (customization), reiteration, self-efficacy, and motivation. For this study, these constructs are commonly used in the study trials compared with other constructs in the literature, such as perceived risk and perceived benefits, incentives, and reimbursement (rewards), which are rarely used in the study trials.

The use of behavioral change constructs in combination with digital technologies in the study trials has revealed their successful application in individual behavioral risk factor modification [7]. The overall desired effect has been found in digital interventions alone (15 studies) when compared with digital plus usual care interventions (10 studies), giving support to out-of-clinic risk factor modification at the personal level [9].

From our results, the use of behavioral change constructs and digital intervention strategies largely relies on patients' self-dependency (low- to moderate-risk CVD patients), and interventions that favored digital technologies were reported for all CVD populations but barely for study trials on stroke and rarely for study trials on angina outpatients. An exception to this reliance was found in studies on medication adherence, which had been successfully self-managed with a digital intervention (SMS text messaging) - this informed advice-based instead of activity-based options for moderate-risk CVD outpatients in risk factor modification prescriptions. The effect

of mobile sensor technology using wearable devices was inconclusive as there was only one study that engaged this digital intervention in its trial (Table S4 in Multimedia Appendix 1).

Clinical Outcomes

There was no clinical benefit for BMI, TG, SBP, DBP, or HbA_{1c} with the use of digital interventions compared with usual care intervention in the study trials. This finding suggests a form of association between clinical factors and unhealthy behavior modification using digital technologies, noting that these clinical factors are the main indicators of unhealthy behavior conditions such as obesity, hypercholesterolemia, hypertension, and diabetes.

However, an exception to these similarities is the significant effect of digital technology intervention on clinical factors such as TC, HDL, and LDL, which could only be inferred by their shared physiological response to regulations in and by healthy behavioral factors such as healthy diet and medication adherence from a lifestyle perspective [45]. This response was less impactful on TGs, which are stored lipids in fatty cells and though considered bad cholesterol like LDL, are less regulated by medication (eg, statin) as compared with diet [45]. Our findings suggest that this exception is not necessarily based on the application of behavioral change techniques or other change-effecting variables. This view is validated by the fact that BMI (an indicator of overweight) and HbA_{1c} (an indicator of excess sugar), which are precursors for obesity and diabetes respectively as CVD risk factors, in association with unhealthy food intake appear to be not modifiable by digital technology interventions. Furthermore, we consider that the modification of LDL (bad cholesterol) by digital intervention might have been because of the positive inverse effect derived from the modification benefits of TC and HDL (good cholesterols) within each study population.

TC, HDL, and LDL modifications are associated with the use of cell phone devices in study trials. Behavior change techniques in TC, HDL, and LDL populations include self-reporting and self-recording of progress, one-on-one support, and persuasion. TC, HDL, and LDL study populations share commonly diagnosed CVDs and digital intervention strategies.

Behavioral Outcomes

PA trials are characterized by smartphones and cell phone devices in a 1:1 ratio. In order of preference, the use of intervention strategies is, first, telerehabilitation and online education, followed by online feedback (tele-support) and telemonitoring, and finally, SMS text messaging support by active coaching (20% of trials). PI (sedentary) trials revealed a higher population mean age, which suggests a close association with comorbidity and immobility among outpatients [46]; therefore, there is a need to tailor digital intervention treatment to patients' level of engagement.

PA and PI have gained modification preferences and digital intervention effectiveness because of active participant engagement in 11 smartphone studies using tele-intervention (audiovisual) strategy, 2 cell phone studies engaging in active coaching (audio) strategy, and 6 cell phone studies using automated SMS text messaging support (text) strategy (in 19 studies). This finding suggests greater effectiveness of smartphones in audiovisual interventions compared with cell phones, indicating that cell phones might have gained usage (in medication adherence study trials) only because of their affordability and ease of use [44,47]. However, both audiovisual support and audio or text support appear to be efficient digital interventions for risk factor modification for PA; however, audio or text support only appears sufficient for PI modification.

Generally, PA (a healthy behavioral factor) has been viewed as a null to PI (sedentary; an unhealthy behavioral factor) effects in maintaining a healthy lifestyle. This view has been disapproved in the literature [48]. However, this disapproval has only been validated in a healthy population prospective study. More evidence is needed to validate this in a CVD population to elucidate the effectiveness of digital technology interventions for PI risk factor modification. Therefore, we consider that the modification of PI, just as in LDL, might have been because of the positive inverse effect of PA modification in the CVD populations reviewed.

Studies reporting effects on diet, alcohol consumption, and smoking share similar characteristics in behavioral change techniques, which include mostly social support and group discussion, followed by self-management, goal setting, follow-up, progress self-reporting and self-recording, and auto-reminders. Social support and group discussion, which are related to online support and online discussion, have been identified as activity-based behavioral change techniques in mental health management for diet, alcohol consumption, and smoking behavior modifications [49]. Interactivity (as a result of social support and group discussion) can, therefore, be affirmed as an effective factor in the behavior change technique of diet, alcohol consumption, and smoking on a digital platform. However, digital interventions for alcohol consumption and smoking behavior change show a weak effect in their modification when compared with conventional CVDs' usual care interventions. There could be several reasons for this—first, social support and group discussions or interaction are less effectively accomplished compared with cell phone device interventions, which have no *smart* facial contact technology features but have gained wider usage in reviewed study trials because of their affordability and availability to participants in both risk factor studies. Second, digital technology interventions, from the trend seen in this study, appear to be effective in healthy behavior modification but less effective in attending to unhealthy behavior modification when compared with usual care: healthy dieting is physiologically linked to lipid regulation in the body [50], a strong basis for clinical factor (TC, HDL, and LDL) modification.

In addition, digital intervention effectiveness in TC, HDL, and LDL, as stated earlier, might also be largely linked to the positive pharmacological effect of medication adherence in study trials. Of the 6 studies on food intake (diet), unhealthy food intake (Figure 14) modification is not favored by digital intervention (P=.47); however, a healthy diet (Figure 13) shows a significant modification effect in favor of digital intervention (P=.004) when compared with usual care. This difference reveals significant alignment and potency of digital intervention

toward healthy behavioral factors than unhealthy behavioral factors. The same was confirmed for PA, a healthy behavioral factor. Healthy behavioral factor (eg, PA, healthy diet, medication adherence) modification using digital technology is supported by findings from Chow et al [18].

Medication adherence outcome from trials in this study was achieved only by the use of cell phones with SMS text messaging support strategy, in line with the findings of Palmer et al [51]. However, the effectiveness of smartphones is inconclusive, as only 1 trial is available in this study—this could be responsible for its limitation in maximizing change technique features, for example, telerehabilitation in medication adherence trials. Cell phones remain the most affordable and available [47] digital devices in medication adherence-targeted interventions compared with other behavioral factor interventions as they cut across all CVD types and engage behavioral change techniques based on cognition such as auto-support, auto-reminders, persuasion (iteration), goal setting, self-management, and customization (personalization).

Trials (Figure 19) that strictly targeted medication adherence outcome only, using an SMS text messaging strategy with a cell phone device, did not show a significant effect (P=.11) when collectively analyzed for digital intervention effectiveness. However, trials (Figure 18) with a similar strategy and device as the former but having multiple clinical and behavioral outcome treatments (analyzed with or without the previous trials) were significantly effective (P=.02) with the use of digital intervention compared with usual care. Non-SMS-administered medication adherence trials (Figure 20) did not favor digital intervention. In summary, these findings suggest the effectiveness of multiple clinical and behavioral outcome treatments when designing digital technology (SMS text messaging) interventions.

A few meta-analyzed results such as for smoking, LDL, BMI, and SBP were limited by high heterogeneity not fully explained (or not explained at all as for alcohol consumption and sedentary lifestyle with low included study counts) by study population or intervention characteristics. However, minor adjustments (exclusion of Chow et al [18], Widmer et al [8], and Redfern et al [33]) in the number of included studies toward increased homogeneity did not show a significant change from the initial treatment effect by either digital intervention or usual care.

The main intervention strategies in this study are automated SMS text messaging support (auto-reminder based on cognition which is largely accessible using cell phones in study trials), a feature supported by Kassavou et al [44]; and then online education and coaching, followed by telerehabilitation and telemonitoring, which were barely represented in analyses that favored digital intervention—representation might be because of limited access to smartphones based on the participants' affordability or level of technological advancement or inclination at the time of the trial. A desirable device is the smartphone because it combines all operability features needed to attain desirable intervention outcomes by identifying behavioral change-specific strategies. However, a major limitation to the use of smartphones by the population age group in the study could be their level of comprehensibility [5].

Limitations

Although this collection of studies is evenly distributed on a global scale, no RCT study has been identified in Africa, where only cost-effective digital health programs have presently gained widespread use [47]. A high proportion of male to female patients would be considered a major limitation of participant inclusion in studies. However, this trend appears to be in resonance with quantitative analyses of CVD gender prevalence in the literature [4] and, therefore, may reflect disease prevalence rather than study design.

This study further reveals gaps in the application of emerging technologies (immersive media, eg, 3D animations and games, an ongoing trial by Gallagher et al [52]; big data technologies, eg, artificial intelligence applications; and user experience) in CVD risk factor modification using evidence-based RCT intervention studies on a digital device platform. Therefore, this study suggests the initiation of cutting-edge research in the field of emerging digital technologies.

Conclusions

This study shows that the use of digital technology interventions did not improve all CVD lifestyle risk factors compared with usual care interventions. Effective digital technology interventions appear to improve healthy behavioral factors (PA, healthy diet) and associated clinical outcomes (TC, HDL, and LDL), and were more potent in multiple outcome treatment (medication adherence plus) but were weak in abating unhealthy behavioral factors (smoking, alcohol intake, and unhealthy food intake) and their outcomes (BMI, BP, and HbA_{1c}).

Cell phones are considered efficient digital devices for use with cognitive intervention strategies and have been most widely studied; however, smartphones may have advantages because of additional interaction features. This study was not able to analyze cutting-edge technology (such as immersive media technologies) as the data do not exist or are not reported. Newer immersive media technologies, therefore, warrant further study. Further RCT research is deemed necessary to consolidate the use of digital technology interventions, especially in CVD risk factors (eg, diabetes), with fewer RCT studies.

Acknowledgments

This is a trans-European cardiovascular (ECME) collaboration project, supervised by the University of the Highlands and Islands, UK. This study was supported by the EU's INTERREG VA Program, managed by the Special EU Programs Body (SEUPB).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Protocol registration link and tables and abbreviations and definition of terms.
[DOC File , 469 KB - mhealth_v9i3e21061_app1.doc]

Multimedia Appendix 2
Study characteristics.
[DOC File , 95 KB - mhealth_v9i3e21061_app2.doc]

References

1. Cardiovascular diseases. World Health Organization. 2019. URL: https://www.who.int/health-topics/cardiovascular-diseases/ [accessed 2021-02-11]
2. Hospital Episode Statistics (HES). NHS Digital. 2017. URL: http://www.content.digital.nhs.uk/hes [accessed 2021-02-11]
3. Health diseases. British Heart Foundation. 2014. URL: https://www.bhf.org.uk/old-starts-with-your-heart/heart-diseases [accessed 2021-02-11]
4. Khot UN, Khot MB, Bajzer CT, Sapp SK, Ohman EM, Brener SJ, et al. Prevalence of conventional risk factors in patients with coronary heart disease. J Am Med Assoc 2003 Aug 20;290(7):898-904. [doi: 10.1001/jama.290.7.898] [Medline: 12928466]
5. Rocca HB, Fleischhacker L, Golubnitschaja O, Heemskerk F, Helms T, Hoedemakers T, et al. Challenges in personalised management of chronic diseases-heart failure as prominent example to advance the care process. EPMA J 2015;7:2 [FREE Full text] [doi: 10.1186/s13167-016-0051-9] [Medline: 26913090]
6. Conner CT. Social problems: the limits and strengths of the social constructionist perspective. Symb Interact 2014 Oct 06. [doi: 10.1002/symb.125]
7. Valente TW. Network interventions. Sci 2012 Jul 06;337(6090):49-53. [doi: 10.1126/science.1217330] [Medline: 22767921]
8. Widmer RJ, Collins NM, Collins CS, West CP, Lerman LO, Lerman A. Digital health interventions for the prevention of cardiovascular disease: a systematic review and meta-analysis. Mayo Clin Proc 2015 Apr;90(4):469-480 [FREE Full text] [doi: 10.1016/j.mayocp.2014.12.026] [Medline: 25841251]
9. Franklin NC, Lavie CJ, Arena RA. Personal health technology: a new era in cardiovascular disease prevention. Postgrad Med 2015 Mar;127(2):150-158 [FREE Full text] [doi: 10.1080/00325481.2015.1015396] [Medline: 25690685]
10. Scotland Government. eHealth strategy 2011-2017 (Revised July 2012 to include a Sixth Strategic Aim). In: The eHealth Strategy for NHS Scotland 2011-2017. Scotland: Scotland Government; 2012.
11. Digital technologies improve remote rehabilitation, medical monitoring for patients. Barcelona Digi Cent Tecnològic. 2019. URL: https://www.sciencedaily.com/releases/2013/11/131104035243.htm [accessed 2019-10-13]
12. Krick T, Huter K, Domhoff D, Schmidt A, Rothgang H, Wolf-Ostermann K. Digital technology and nursing care: a scoping review on acceptance, effectiveness and efficiency studies of informal and formal care technologies. BMC Health Serv Res 2019 Jun 20;19(1):400 [FREE Full text] [doi: 10.1186/s12913-019-4238-3] [Medline: 31221133]
13. Charles A, Buck D, Evans H, McKenna H. King's Fund Webcast: what's in store for health and care? King's Fund Webcast, 2019. URL: https://www.kingsfund.org.uk/events/whats-store-health-and-care-2020 [accessed 2021-02-11]
14. 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]
15. Akinosun AS, Polson R, Grindle M. Digital technologies risk factors modification in patients with cardiovascular disease: a systematic review and meta-analysis. Prospero. 2019. URL: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019139801 [accessed 2021-02-11]
16. World Health Organization, global strategy on digital health 2020-2024 in global strategy on digital health. 2019. URL: https://extranet.who.int/dataform/upload/surveys/183439/files/Draft [accessed 2021-02-11]
17. Akhu-Zaheya LM, Shiyab WY. The effect of short message system (SMS) reminder on adherence to a healthy diet, medication, and cessation of smoking among adult patients with cardiovascular diseases. Int J Med Inform 2017 Feb;98:65-75. [doi: 10.1016/j.ijmedinf.2016.12.003] [Medline: 28034414]
18. Chow CK, Redfern J, Hillis GS, Thakkar J, Santo K, Hackett ML, et al. Effect of lifestyle-focused text messaging on risk factor modification in patients with coronary heart disease: a randomized clinical trial. J Am Med Assoc 2015;314(12):1255-1263. [doi: 10.1001/jama.2015.10945] [Medline: 26393848]
19. Pfaeffli Dale L, Whittaker R, Jiang Y, Stewart R, Rolleston A, Maddison R. Text message and internet support for coronary heart disease self-management: results from the text4heart randomized controlled trial. J Med Internet Res 2015 Oct 21;17(10):e237 [FREE Full text] [doi: 10.2196/jmir.4944] [Medline: 26490012]
20. Devi R, Powell J, Singh S. A web-based program improves physical activity outcomes in a primary care angina population: randomized controlled trial. J Med Internet Res 2014 Sep 12;16(9):e186 [FREE Full text] [doi: 10.2196/jmir.3340] [Medline: 25217464]

21. Frederix I, Van Driessche N, Hansen D, Berger J, Bonne K, Alders T, et al. Increasing the medium-term clinical benefits of hospital-based cardiac rehabilitation by physical activity telemonitoring in coronary artery disease patients. Eur J Prev Cardiol 2015 Feb;22(2):150-158. [doi: 10.1177/2047487313514018] [Medline: 24249840]
22. Hawkes AL, Patrao TA, Atherton J, Ware RS, Taylor CB, O'Neil A, et al. Effect of a telephone-delivered coronary heart disease secondary prevention program (proactive heart) on quality of life and health behaviours: primary outcomes of a randomised controlled trial. Int J Behav Med 2013 Sep;20(3):413-424. [doi: 10.1007/s12529-012-9250-5] [Medline: 23012159]
23. Johnston N, Bodegard J, Jerström S, Åkesson J, Brorsson H, Alfredsson J, et al. Effects of interactive patient smartphone support app on drug adherence and lifestyle changes in myocardial infarction patients: A randomized study. Am Heart J 2016 Aug;178:85-94 [FREE Full text] [doi: 10.1016/j.ahj.2016.05.005] [Medline: 27502855]
24. Kamal AK, Shaikh Q, Pasha O, Azam I, Islam M, Memon AA, et al. A randomized controlled behavioral intervention trial to improve medication adherence in adult stroke patients with prescription tailored Short Messaging Service (SMS)-SMS4Stroke study. BMC Neurol 2015 Oct 21;15:212 [FREE Full text] [doi: 10.1186/s12883-015-0471-5] [Medline: 26486857]
25. Khonsari S, Subramanian P, Chinna K, Latif LA, Ling LW, Gholami O. Effect of a reminder system using an automated short message service on medication adherence following acute coronary syndrome. Eur J Cardiovasc Nurs 2015 Apr;14(2):170-179. [doi: 10.1177/1474515114521910] [Medline: 24491349]
26. Kraal JJ, Peek N, Van den Akker-Van Marle ME, Kemps HM. Effects of home-based training with telemonitoring guidance in low to moderate risk patients entering cardiac rehabilitation: short-term results of the FIT@Home study. Eur J Prev Cardiol 2014 Nov;21(2 Suppl):26-31. [doi: 10.1177/2047487314552606] [Medline: 25354951]
27. Lear SA, Singer J, Banner-Lukaris D, Horvat D, Park JE, Bates J, et al. Randomized trial of a virtual cardiac rehabilitation program delivered at a distance via the internet. Circ Cardiovasc Qual Outcomes 2014 Nov;7(6):952-959. [doi: 10.1161/CIRCOUTCOMES.114.001230] [Medline: 25271050]
28. Maddison R, Pfaeffli L, Whittaker R, Stewart R, Kerr A, Jiang Y, et al. A mobile phone intervention increases physical activity in people with cardiovascular disease: results from the HEART randomized controlled trial. Eur J Prev Cardiol 2015 Jun;22(6):701-709. [doi: 10.1177/2047487314535076] [Medline: 24817694]
29. Ögren J, Irewall AL, Söderström L, Mooe T. Long-term, telephone-based follow-up after stroke and TIA improves risk factors: 36-month results from the randomized controlled NAILED stroke risk factor trial. BMC Neurol 2018 Sep 21;18(1):153 [FREE Full text] [doi: 10.1186/s12883-018-1158-5] [Medline: 30241499]
30. Pandey A, Choudhry N. Text message reminders to address medication non-adherence in post-mi patients: a one year intervention study. Can J Cardiol 2014 Oct;30(10):S179. [doi: 10.1016/j.cjca.2014.07.280]
31. Park LG, Howie-Esquivel J, Chung ML, Dracup K. A text messaging intervention to promote medication adherence for patients with coronary heart disease: a randomized controlled trial. Patient Educ Couns 2014 Feb;94(2):261-268. [doi: 10.1016/j.pec.2013.10.027] [Medline: 24321403]
32. Quilici J, Fugon L, Beguin S, Morange PE, Bonnet JL, Alessi MC, et al. Effect of motivational mobile phone short message service on aspirin adherence after coronary stenting for acute coronary syndrome. Int J Cardiol 2013 Sep 20;168(1):568-569. [doi: 10.1016/j.ijcard.2013.01.252] [Medline: 23462636]
33. Redfern J, Briffa T, Ellis E, Freedman SB. Choice of secondary prevention improves risk factors after acute coronary syndrome: 1-year follow-up of the CHOICE (Choice of Health Options In prevention of Cardiovascular Events) randomised controlled trial. Heart 2009 Mar;95(6):468-475. [doi: 10.1136/hrt.2008.150870] [Medline: 18801781]
34. Reid RD, Morrin LI, Beaton LJ, Papadakis S, Kocourek J, McDonnell L, et al. Randomized trial of an internet-based computer-tailored expert system for physical activity in patients with heart disease. Eur J Prev Cardiol 2012 Dec;19(6):1357-1364. [doi: 10.1177/1741826711422988] [Medline: 21903744]
35. Southard BH, Southard DR, Nuckolls J. Clinical trial of an internet-based case management system for secondary prevention of heart disease. J Cardiopulm Rehabil 2003;23(5):341-348. [doi: 10.1097/00008483-200309000-00003] [Medline: 14512778]
36. Tiede M, Dwinger S, Herbarth L, Härter M, Dirmaier J. Long-term effectiveness of telephone-based health coaching for heart failure patients: a post-only randomised controlled trial. J Telemed Telecare 2017 Sep;23(8):716-724. [doi: 10.1177/1357633X16668436] [Medline: 27605214]
37. Vale MJ, Jelinek MV, Best JD, Santamaria JD. Coaching patients with coronary heart disease to achieve the target cholesterol: a method to bridge the gap between evidence-based medicine and the "real world"--randomized controlled trial. J Clin Epidemiol 2002 Mar;55(3):245-252. [doi: 10.1016/s0895-4356(01)00460-7] [Medline: 11864795]
38. Vernooij JW, Kaasjager HA, van der Graaf Y, Wierdsma J, Grandjean HM, Hovens MM, SMARTStudy Group. Internet based vascular risk factor management for patients with clinically manifest vascular disease: randomised controlled trial. BMJ 2012 Jun 12;344:e3750 [FREE Full text] [doi: 10.1136/bmj.e3750] [Medline: 22692651]
39. Wan LH, Zhang XP, Mo MM, Xiong XN, Ou CL, You LM, et al. Effectiveness of goal-setting telephone follow-up on health behaviors of patients with ischemic stroke: a randomized controlled trial. J Stroke Cerebrovasc Dis 2016 Sep;25(9):2259-2270. [doi: 10.1016/j.jstrokecerebrovasdis.2016.05.010] [Medline: 27371106]

40. Widmer RJ, Allison TG, Lennon R, Lopez-Jimenez F, Lerman LO, Lerman A. Digital health intervention during cardiac rehabilitation: a randomized controlled trial. Am Heart J 2017 Jun;188:65-72. [doi: 10.1016/j.ahj.2017.02.016] [Medline: 28577682]
41. Zheng X, Spatz ES, Bai X, Huo X, Ding Q, Horak P, et al. Effect of text messaging on risk factor management in patients with coronary heart disease: the CHAT randomized clinical trial. Circ Cardiovasc Qual Outcomes 2019 Apr;12(4):e005616. [doi: 10.1161/CIRCOUTCOMES.119.005616] [Medline: 30998400]
42. Kraal JJ, Peek N, van den Akker-Van Marle ME, Kemps HMC. Effects and costs of home-based training with telemonitoring guidance in low to moderate risk patients entering cardiac rehabilitation: the FIT@Home study. BMC Cardiovasc Disord 2013 Oct 08;13:82 [FREE Full text] [doi: 10.1186/1471-2261-13-82] [Medline: 24103384]
43. Vale MJ, Jelinek MV, Best JD, Santamaria JD. Coaching patients with coronary heart disease to achieve the target cholesterol: a method to bridge the gap between evidence-based medicine and the. J Clin Epidemiol 2002 Mar;55(3):245-252. [doi: 10.1016/s0895-4356(01)00460-7] [Medline: 11864795]
44. Kassavou A, Houghton V, Edwards S, Brimicombe J, Sutton S. Development and piloting of a highly tailored digital intervention to support adherence to antihypertensive medications as an adjunct to primary care consultations. BMJ Open 2019 Jan 06;9(1):e024121 [FREE Full text] [doi: 10.1136/bmjopen-2018-024121] [Medline: 30613027]
45. Lower your cholesterol. NHS. 2019. URL: https://www.nhs.uk/live-well/healthy-body/lower-your-cholesterol/ [accessed 2019-10-18]
46. Figueiró TH, Arins GCB, Santos CESD, Cembranel F, Medeiros PAD, d'Orsi E, et al. Association of objectively measured sedentary behavior and physical activity with cardiometabolic risk markers in older adults. PLoS One 2019 Jan 18;14(1):e0210861 [FREE Full text] [doi: 10.1371/journal.pone.0210861] [Medline: 30657795]
47. Healthcare and economic growth in Africa: digital health. UNECA, GBCHealth and Aliko Dangote Foundation. 2019. URL: http://gbchealth.org/healthcare-and-economic-growth-in-africa-digital-health/ [accessed 2021-02-11]
48. Katzmarzyk PT. Physical activity, sedentary behavior, and health: paradigm paralysis or paradigm shift? Diabetes 2010 Nov;59(11):2717-2725 [FREE Full text] [doi: 10.2337/db10-0822] [Medline: 20980470]
49. National Institute for Health and Care Excellence. Behaviour change: digital and mobile health interventions. NICE. 2020. URL: https://www.nice.org.uk/guidance/gid-ng10101/documents/final-scope [accessed 2021-02-11]
50. Udenigwe CC, Rouvinen-Watt K. The role of food peptides in lipid metabolism during dyslipidemia and associated health conditions. Int J Mol Sci 2015 Apr 24;16(5):9303-9313 [FREE Full text] [doi: 10.3390/ijms16059303] [Medline: 25918936]
51. Palmer MJ, Barnard S, Perel P, Free C. Mobile phone-based interventions for improving adherence to medication prescribed for the primary prevention of cardiovascular disease in adults. Cochrane Database Syst Rev 2018 Jun 22;6:CD012675 [FREE Full text] [doi: 10.1002/14651858.CD012675.pub2] [Medline: 29932455]
52. Gallagher R, Chow C, Parker H, Neubeck L, Celermajer D, Redfern J, et al. Design and rationale of the MyHeartMate study: a randomised controlled trial of a game-based app to promote behaviour change in patients with cardiovascular disease. BMJ Open 2019 May 14;9(5):e024269 [FREE Full text] [doi: 10.1136/bmjopen-2018-024269] [Medline: 31092643]

Abbreviations

BP: blood pressure
CHD: coronary heart disease
CVD: cardiovascular disease
DBP: diastolic blood pressure
HbA_{1c}: Hemoglobin A_{1c} blood sugar
HDL: high-density lipoprotein
LDL: low-density lipoprotein
MD: mean difference
PA: physical activity
PI: physical inactivity
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analysis
RCT: randomized controlled trial
RR: risk ratio
SBP: systolic blood pressure
SMD: standard mean difference
TC: total cholesterol
TG: triglyceride

Edited by L Buis; submitted 05.06.20; peer-reviewed by L Park, R Lundin, J Li, J Van den Eynde, N Mohammad Gholi Mezerji; comments to author 14.09.20; revised version received 08.11.20; accepted 01.12.20; published 03.03.21.

Please cite as:
Akinosun AS, Polson R, Diaz - Skeete Y, De Kock JH, Carragher L, Leslie S, Grindle M, Gorely T
Digital Technology Interventions for Risk Factor Modification in Patients With Cardiovascular Disease: Systematic Review and Meta-analysis
JMIR Mhealth Uhealth 2021;9(3):e21061
URL: https://mhealth.jmir.org/2021/3/e21061
doi:10.2196/21061
PMID:33656444

Review

Factors Affecting the Quality of Person-Generated Wearable Device Data and Associated Challenges: Rapid Systematic Review

Sylvia Cho[1], MHS, MA; Ipek Ensari[2], PhD; Chunhua Weng[1], PhD; Michael G Kahn[3], PhD, MD; Karthik Natarajan[1,2], PhD

[1]Department of Biomedical informatics, Columbia University, New York, NY, United States

[2]Data Science Institute, Columbia University, New York, NY, United States

[3]Department of Pediatrics, University of Colorado Anschutz Medical Campus, Denver, CO, United States

Corresponding Author:
Sylvia Cho, MHS, MA
Department of Biomedical informatics
Columbia University
622 West 168th Street PH20
New York, NY, 10032
United States
Phone: 1 212 305 5334
Email: sc3901@cumc.columbia.edu

Abstract

Background: There is increasing interest in reusing person-generated wearable device data for research purposes, which raises concerns about data quality. However, the amount of literature on data quality challenges, specifically those for person-generated wearable device data, is sparse.

Objective: This study aims to systematically review the literature on factors affecting the quality of person-generated wearable device data and their associated intrinsic data quality challenges for research.

Methods: The literature was searched in the PubMed, Association for Computing Machinery, Institute of Electrical and Electronics Engineers, and Google Scholar databases by using search terms related to wearable devices and data quality. By using PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, studies were reviewed to identify factors affecting the quality of wearable device data. Studies were eligible if they included content on the data quality of wearable devices, such as fitness trackers and sleep monitors. Both research-grade and consumer-grade wearable devices were included in the review. Relevant content was annotated and iteratively categorized into semantically similar factors until a consensus was reached. If any data quality challenges were mentioned in the study, those contents were extracted and categorized as well.

Results: A total of 19 papers were included in this review. We identified three high-level factors that affect data quality—device- and technical-related factors, user-related factors, and data governance-related factors. Device- and technical-related factors include problems with hardware, software, and the connectivity of the device; user-related factors include device nonwear and user error; and data governance-related factors include a lack of standardization. The identified factors can potentially lead to intrinsic data quality challenges, such as incomplete, incorrect, and heterogeneous data. Although missing and incorrect data are widely known data quality challenges for wearable devices, the heterogeneity of data is another aspect of data quality that should be considered for wearable devices. Heterogeneity in wearable device data exists at three levels: heterogeneity in data generated by a single person using a single device (within-person heterogeneity); heterogeneity in data generated by multiple people who use the same brand, model, and version of a device (between-person heterogeneity); and heterogeneity in data generated from multiple people using different devices (between-person heterogeneity), which would apply especially to data collected under a bring-your-own-device policy.

Conclusions: Our study identifies potential intrinsic data quality challenges that could occur when analyzing wearable device data for research and three major contributing factors for these challenges. As poor data quality can compromise the reliability and accuracy of research results, further investigation is needed on how to address the data quality challenges of wearable devices.

(JMIR Mhealth Uhealth 2021;9(3):e20738) doi:10.2196/20738

KEYWORDS

patient generated health data; data accuracy; data quality; wearable device; fitness trackers; mobile phone

Introduction

Emerging Biomedical Data—Person-Generated Wearable Device Data

With the recent movement toward people (patient)-centered care and the widespread routine use of devices/technologies, person-generated health data (PGHD) have emerged as a promising data source for biomedical research [1]. A survey conducted in 2019 reported that 38% of Americans currently use technologies such as mobile apps or wearables to track their health data, and 28% have used them in the past [2]. Examples of PGHD include data collected passively through sensors, such as step count, heart rate, and sleep quality; data entered directly by people, such as diet, stress levels, and quality of life; and social or financial information that is not specifically health related but could potentially provide health-related insights [3]. Among the different PGHD, data generated through wearable devices are unique in that they are passively, continuously, and objectively collected in free-living conditions; such data are different from those generated through other technologies that require the manual input of data (eg, dietary tracking mobile apps) [4-7]. Therefore, person-generated wearable device data are becoming a valuable resource for biomedical researchers to provide a more comprehensive picture of the health of individuals and populations.

Use of Person-Generated Wearable Device Data for Research Purposes

There are two ways to use wearable device data for research purposes. Typically, researchers collect wearable device data for a specific research by recruiting eligible participants and asking them to use the device for a certain period. For example, Lim et al [8] issued Fitbit devices to 233 participants and asked them to use the device for 5 days. Collecting data with this traditional method can be beneficial in that people can collect data that fits their needs, but it can be costly to recruit and follow a large number of participants for an extended period.

Researchers can also reuse existing data, which is a timely and cost-effective way to conduct research. Previous studies have used existing wearable device data collected for other research studies for their own research [8,9]. For example, McDonald et al [9] used a data set collected as part of the SingHEART/Biobank study to investigate the association between sleep and body mass index. In addition, Cheung et al [10] used data collected from a study by Burg et al [11] to develop a novel methodology to reduce the dimension of data while maintaining core information.

More recently, real-world wearable device data collected through routine use of devices have been reused for research purposes [7,12,13]. For example, the All of Us research program, which is the precision medicine initiative launched by the National Institutes of Health (NIH), initiated a Fitbit Bring-Your-Own-Device project, which allows participants to connect their Fitbit account to share data, such as physical activity, sleep, and heart rate [14]. In addition, multiple studies have shown the potential of routinely collected wearable device data for use in large-scale longitudinal multinational studies. Menai et al [15] used Withings Pulse activity tracker data of 9238 adults from 37 countries collected from 2009 to 2013 to examine the association between step counts and blood pressure. Kim et al [16] used data of more than 50,000 individuals from 185 countries collected over a month, with nearly 17 million measurements generated by Nokia Health Wireless blood pressure monitors to characterize blood pressure variability. These studies underscore the potential secondary uses of person-generated wearable device data for generating health insights from large real-world population that might not have been possible using traditional methods of data collection. Furthermore, the studies demonstrate how wearable device data add value by expanding the scope of biomedical research that can be conducted, which would not have been feasible if relying on electronic health record (EHR) data alone.

Data Quality Challenges in the Use of Person-Generated Wearable Device Data

Data used in research studies, even data originally collected to support research, may not meet the ideal level of quality [13,17,18]. For instance, data collected daily through consumer wearables are meant to be used for routine use of devices rather than for research. Therefore, although the quality of collected data may be sufficient for an individual's health management, it may be insufficient for research purposes. Hicks et al [19] presented the best practices for reusing large-scale consumer wearable device data that were collected through routine use. The study describes challenges with data quality, such as missing data or inaccuracy of sensor data, as these data are collected from individuals through their daily use of wearables (not through a research study). Thus, as recommended for the use of any data set, the study recommends assessing the quality of wearable device data set before conducting research. Once the research question and data set to be analyzed are identified, it is important to assess its fitness-for-use to ensure that it would produce valid analytical results that answer the research question [19].

There have been previous efforts to understand the data quality challenges for wearable device data. For example, Codella et al [7] identified the data quality dimensions that influence the analysis of PGHD. The concerns and expectations of PGHD stakeholders were identified through a literature review and mapped to the relevant data quality dimensions of an established framework [7]. However, the review does not systematically provide the steps of how they screened and selected the literature and what information they extracted within the studies. Another systematic review by Abdolkhani et al [20] identified factors influencing the quality of medical wearable device data and their corresponding dimensions from the literature. However, this review did not include literature on data from nonmedically approved wearables (eg, consumer wearable devices). As such, there is a research gap in understanding data quality challenges

that arise from consumer wearables, specifically those from passively collected data, as there might be unique quality challenges associated with these types of data.

Objectives

While assessing data quality, having a full understanding of the types of data quality challenges and the factors associated with them can be useful in implementing additional analytic procedures to ameliorate potential negative impacts or false conclusions. However, one of the barriers is that there is a lack of studies investigating the data quality challenges of wearable device data specifically for research purposes. Therefore, this study aims to (1) identify factors influencing the quality of person-generated wearable device data and potential intrinsic data quality challenges (data quality in its own right or, in other words, data quality challenges inherent to the data itself) for research, and (2) discuss implications for the appropriate use of person-generated wearable device data for research purposes based on the findings [21].

Methods

Data Sources and Search Strategy

We performed a rapid review following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) guidelines. The literature search was conducted in four scholarly databases (PubMed, Association for Computing Machinery [ACM] Digital Library, Institute of Electrical and Electronics Engineers [IEEE], and Google Scholar) in June 2019. In PubMed, we used a combination of MeSH terms and keywords related to wearable devices and data quality. Terms related to mobile health were not searched because they include mobile apps or telemedicine, although the scope of this review focused specifically on passively collected data through wearable devices. The search results were limited to studies published within the past 5 years, studies conducted with human species and studies written in English language. The search was limited to 2014 onward because the characteristics of devices may change with advances in technologies, and this may result in changes in data quality challenges. Thus, the search was focused on recent publications using the year with the largest increase in the emergence of new consumer fitness trackers as a heuristic cutoff for determining recent studies [12]. The publications were sorted by best match, which is appropriate for searching studies that meet the informational needs on a topic [22].

In the ACM Digital Library and IEEE Xplore Digital Library, we used a query that combined search terms related to data quality and wearable devices. The search results were limited to studies published since 2014. To complement the search results from the 3 scholarly databases, we performed an additional literature search on Google Scholar. In total, 4 searches were conducted using different queries. The search excluded patents and citations, examined studies published since 2014, and sorted the results by relevance. Although all of the search results were reviewed for other scholarly databases, only the first 100 results for each of the 4 queries in Google Scholar were reviewed. To prevent the filter bubble effect, which customizes search results based on the search history of users, Google accounts were logged out when conducting the literature search [23]. The full query used in each database can be found in Table S1 in Multimedia Appendix 1.

Literature Selection

Inclusion criteria were as follows: (1) papers that contained content on the data quality of wearable devices or sensor data; (2) papers that demonstrated the scope of wearable devices, including devices such as fitness trackers, sleep monitors, continuous glucose monitors, and remote blood pressure trackers; (3) papers on research-grade and consumer-grade devices; and (4) not only peer-reviewed studies, but also conference proceedings and book chapters to expand the search space.

Although smartphones can passively collect health data, studies that exclusively focused on smartphones were excluded, as they are not worn on the body. In addition, as we were interested in passively collected person-generated wearable device data being used for research, studies were excluded if (1) the study was on wearable device data that were generated by providers in a clinical setting (eg, device being used for clinician or surgical training), (2) the study was on wearable device data being used for clinical care of patients, and (3) the study was on data that were manually recorded (eg, food logging by user). Device validation studies such as testing the accuracy, reliability, or validity of the device were also excluded, as those studies were about testing the accuracy of the device rather than conducting analyses on data.

One reviewer (SC) screened the retrieved literature based on the title and abstract. After filtering based on titles and abstracts, the full text of the remaining studies was reviewed based on the same selection criteria by two reviewers (SC and KN). The reviewers discussed any discrepancies to reach a consensus on the final set of studies. The literature selection process was conducted using Covidence (Veritas Health Innovation), which is a web-based systematic review production tool.

Data Extraction and Categorization

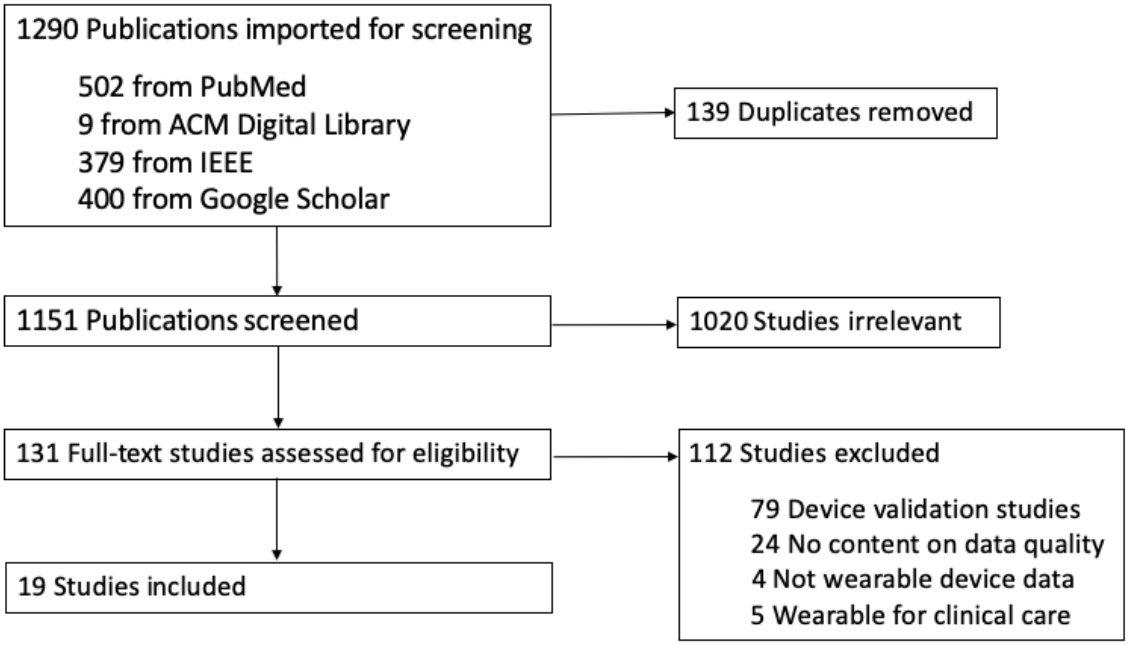
Overall, two reviewers (SC and KN) examined the papers to extract sentences about the factors affecting data quality. Although our focus was on wearable device data, sentences that apply to both mobile app and wearable device data were extracted as long as the content did not exclusively apply to mobile app data. The reviewers extracted the sentences and annotated the relevant factors. In addition, intrinsic data quality challenges associated with those factors were extracted if any were mentioned. Microsoft Excel was used to manage qualitative data. Codes were assigned to phrases that indicated factors influencing data quality by 1 reviewer (SC). Coded concepts were reviewed, and semantically similar concepts were consolidated into the same category. The categories were iteratively refined to derive core categories. The categories were then iteratively reviewed by domain experts (one data quality expert [KN] and one wearable device expert [IE]) to refine and validate the results. Domain experts commented on whether they agreed with the categorization and names used for each category. The discussion continued until a consensus among the reviewers and domain experts was reached.

Results

Literature Search and Selection Results

A total of 1290 publications were retrieved for screening. Among the retrieved publications, 139 duplicates were removed, leaving 1151 unique publications to be screened by title and abstract. The screening of titles and abstracts resulted in 131 studies after removing 1020 publications that did not meet the eligibility criteria. The full texts of the remaining 131 publications were reviewed. After removing 112 irrelevant publications, 19 studies remained. The literature selection process is depicted in Figure 1, and a summary of the included studies can be found in Table S2 in Multimedia Appendix 1.

Figure 1. Flow diagram of the literature selection process. ACM: Association for Computing Machinery; IEEE: Institute of Electrical and Electronics Engineers.



Data Extraction and Categorization Results

Some extracted sentences were specifically related to wearable device data. For instance, sentences within a study by Wright et al [24] describe the challenges associated with using consumer fitness trackers in biomedical research:

> *The algorithms used in consumer physical activity monitors to determine steps taken, distance traveled, and energy expenditure are typically not shared with researchers due to proprietary concerns.*

On the other hand, there were sentences that could apply to both wearable devices and mobile apps. For example, Bietz et al [25] examined data quality challenges of routine use of devices data and explicitly stated the challenges that researchers face:

> *Researchers also reported being concerned with the kinds of data they may get from companies, including the lack of standardization, potential problems with proprietary algorithms, and that most of the consumer-level health devices have not gone through a validation process.*

Not all concerns regarding wearable device data were extracted from these studies. For example, Bietz et al [25] mentioned selection bias, which was not extracted, as we believe that bias is not an intrinsic data quality challenge but is a byproduct of data quality and a universal challenge to research design:

> *A related concern is the potential bias in PGHD that derives from who uses personal health devices and who does not.*

After 5 iterations of categorizing the factors influencing data quality with domain experts, 3 broad categories emerged, which are summarized in Textbox 1. The mappings between the factors and the intrinsic data quality challenges are presented in Table 1.

Textbox 1. Factors influencing data quality and the themes identified in selected literature.

Device- and technical-related factors

- Hardware issues [26-28]
  - Malfunction [26,29-32]
  - Quality of sensor [3,7,24,32-34]
  - Sensor degradation over time [27]
  - Device update makes older models outdated [24]
  - Limited storage space [32]
- Software issues [24,25,27,29,34,35]
  - Quality (accuracy) of algorithm [7,31,33]
  - Proprietary algorithm or system [25,27,29,35]
  - Wearable device companies change and update their algorithms [24]
  - Software updates may change settings to default setting or affect data [34]
- Network and Bluetooth issues [29-31,34,36]
  - Lost satellite connection [29,30,32,34,36]
  - Delay and error in synchronization and data upload [29,30,34,36]

User-related factors

- User nonwear [7,24,26,30,33,34,36]
  - Forget to wear [26,33]
  - Nonwear during battery charging [7,24,30,34,36]
  - User's health condition prevents device use [30]
  - Discomfort of wearing the device [7,24]
  - Unsatisfied with the appearance of device [30]
  - User's lifestyle or not wearing for certain everyday activities [30]
  - Concerns over privacy and security of data [30]
  - Poor usability experience [30]
- User error [27,29-31,33,34,37]
  - Device not synced by users [29]
  - Poor calibration of the device [37]
  - Quality of skin contact [34]
  - Misplacement of device on the body [24,27,34]

Data governance-related factors

- Lack of standardization [3,7,25,33,34,38]
  - No industry standards for data formats, range of values, and sample rates [34,38-40]
  - Different devices use different algorithms for the same variable [3,7,38]
  - Different type or placement of sensors on the body for the same variable [37]
  - Different data definition for the same variable [7,33]

Table 1. Mappings between factors and intrinsic data quality challenges.

Factors influencing DQ[a]	Intrinsic DQ challenges		
	Completeness	Correctness	Heterogeneity
Device- and technical-related factors			
Hardware issues [26,27]	✓[b]	✓	—[c]
Software issues [24,25,27,29,34,35]	—	✓	✓
Network and Bluetooth issues [29-31,34,36]	✓	—	—
User-related factors			
User nonwear [7,24,26,30,33,34,36]	✓	—	—
User error [27,29-31,33,34,37]	✓	✓	—
Data governance-related factors			
Lack of standardization [3,7,25,33,34,38]	—	—	✓

[a]DQ: data quality.

[b]This indicates that the data quality challenge is associated with the factor according to the studies included in the review.

[c]Not available. This indicates that the data quality challenge was not particularly mentioned in studies as an associated challenge of the factor.

Factors Affecting the Quality of Person-Generated Wearable Device Data

Device- and Technical-Related Factors

Device- and technical-related factors consist of issues related to (1) hardware, (2) software, and (3) network and Bluetooth. Issues related to hardware include sensor malfunction [26,29-32], the quality of sensors [3,7,24,32-34], and sensor degradation over time [27]. For instance, companies continuously upgrade their devices, which means that older models are outdated and may no longer be supported by the company [24]. This may affect studies that are interested in longitudinal data, as discontinued device support may lead to incomplete data [24].

There are several issues with software or algorithms used to interpret raw sensor data [24,25,27,29,34,35]. One major issue is that consumer wearables use proprietary algorithms for their devices [25,27,29,35]. Thus, it is difficult to know if or when consumer wearable companies change and update their algorithms [24]. The lack of transparency regarding the timing and impact of software change can impact data consistency between participants who have data from different periods and also between data from the same participant collected longitudinally [24].

Network and Bluetooth problems can also affect the data quality of wearable devices. Lack of wireless signals or lost satellite connections can cause errors and delays in capturing, synchronizing, and uploading the data [29,30,34,36]. In addition, the location tracking function might stop working when the user is in a building with poor satellite connection, which could lead to missing data problems [30].

User-Related Factors

A primary user-related factor is not wearing the device (nonwear time) [7,24,26,30,33,34,36]. Missing data that occur from nonwear is a major limitation to the accuracy of estimates derived from wearables because the pattern of missingness in these instances is often not at random (ie, missing not at random), which has implications for inferences that can be made based on these estimates [41,42]. Another user-related factor is incorrect use by users. For instance, researchers conducting time-sensitive studies should keep in mind that automatic time zone updates may fail, and users may forget to manually update or synchronize their time zone when traveling [29].

Data Governance-Related Factors

Data standard is an essential deliverable of data governance that can not only affect the comparability between data systems but can also influence the researcher's ability to make reliable inferences from data [43]. However, wearable device data, more specifically consumer-grade wearables, are rarely standardized to interoperate with clinical systems, as such devices are developed for consumer use rather than research or clinical practice [44]. Lack of standardization can cause significant heterogeneity across data from different device brands (eg, Fitbit vs Garmin) or different models within the same brand (eg, Fitbit Charge 3 and Fitbit Inspire) and more broadly across individuals and different clinical centers. As a result, it might be difficult for researchers to integrate data sets and make a direct comparison between the analysis results from different device data [3,7,25,33].

Intrinsic Data Quality Challenges of Person-Generated Wearable Device Data

One of the goals of this study was to identify potential data quality challenges when reusing data from the routine use of devices for research purposes. However, because of the lack of literature on the reuse of wearable device data, data quality challenges for research in general have been investigated. As a result of the review, three intrinsic data quality challenges were identified—completeness, correctness, and heterogeneity. Missing data were indicated as challenges occurring because of device malfunction, lost satellite connection or synchronization error, users not wearing the device, and devices unstably contacting the skin [7,26,30,34,36]. Incorrect data, which were more frequently stated as inaccurate data in studies,

was another potential data quality challenge [26,27,33,35]. Poor sensor quality, the unknown limitations of proprietary algorithms, or user errors such as incorrect device placement can all contribute to incorrect data [26,33]. Another problem is the potential heterogeneity across data sources, which can lead to difficulty in intra- and intersubject comparisons [25,35,38]. This is because (1) companies do not always reveal whether or when they update their device algorithms or whether or when the users install the provided software updates, and (2) different devices may use different algorithms or data definitions for the same variable [25,35,38]. The focus of this study was on intrinsic data quality challenges, which are challenges on the data in its own right [21]. Thus, challenges extrinsic to data such as data accessibility, security, and privacy were not included.

Discussion

Principal Findings

Device- and technical-related, user-related, and data governance-related factors were identified as factors that influence the quality of wearable device data. These factors can potentially affect 3 intrinsic data quality challenges: completeness, correctness, and heterogeneity of data. Of note, the factors identified in this review are inherent to the characteristics of wearable device data as opposed to factors that could occur while processing the data, such as factors in extract, transform, and load (ETL) processes [45]. Researchers conducting multicenter studies should keep in mind that converting their wearable device data by using a common data model may induce additional errors during ETL processes [46]. Factors associated with data quality problems were classified into 3 main categories; however, the authors realized that the identified factors were highly connected to each other, and thus, the categorization could be subjective. For example, limited battery life is a device-related feature, but as a low battery level could make the user take off the device to charge the device, it was classified as a user-related factor. In addition, the proprietary algorithm of devices can be a data governance-related factor as proprietary algorithms lead to heterogeneity in multidevice data due to lack of data standards. However, the proprietary algorithm of devices was classified as a device-related factor because algorithms are part of the device and can produce data heterogeneity in single-device data as well. Despite the subjective nature of this work, three researchers iteratively refined the categories until a consensus was reached. As this is an early attempt to investigate data quality challenges for wearable device data, the authors expect this categorization to be refined in the future as researchers start to apply this framework while assessing data quality.

Implications and Recommendations for Researchers

Summary of Recommendations for Researchers

Our study results indicate that a multitude of intrinsic data quality challenges exist for person-generated wearable device data, and we summarize the factors that underlie these challenges. We report completeness, correctness, and heterogeneity of data as the 3 primary concerns for researchers looking to conduct research using data from wearable devices. The implications and recommendations provided in this section are derived from the authors' domain expertise and are based on existing literature both within and outside this review. A summary of the recommendations is presented in Textbox 2.

Textbox 2. Summary of intrinsic data quality challenges and recommendations for researchers.

Completeness

- Report the definition of completeness used in research studies.
- Best practices on fitness-for-use measures for data completeness should be investigated.

Correctness

- Community effort to create a knowledge base of data quality rules is needed.
- Identify methods or external data sources that would help researchers retrospectively assess the plausibility of their data set.

Heterogeneity

- Data providers should collect metadata on which device brand, model, and software version the data are generated from.
- Researchers should check these metadata before conducting analyses and report it when publishing study results.

Data Completeness

Completeness is one of the major data quality challenges for wearable device data, mainly because users do not wear the device. Completeness is also a complex challenge, as various considerations need to be made by researchers to assess it. First, researchers need to determine how they would distinguish between true inactivity and device nonwear. This is especially the case for step count data, as missing data are unique in that they could appear as null values (eg, because of error in the device) or appear as zeros if the device is not worn. This is a challenge, as the cause of zero values (eg, nonwear, sedentary behavior, connectivity issue) is typically not documented, especially if the device is routinely used in daily lives. Previous studies have defined nonwear time with various thresholds for inactivity (zero count of activity) periods ranging from 10 to 60 minutes [47,48]. As different definitions of nonwear time may significantly change the total wear time per day and analysis results, reporting what threshold was used would be an important step for researchers [47].

In addition, there are multiple measures to consider when assessing data completeness among which one is valid day—a day with sufficient data that can be kept for analyses [49]. Tang et al [49] proposed three heuristic criteria for valid days: (1) minimum step count (eg, a day is valid if the daily step count is greater than 500), (2) the minimum count of hours with data (eg, a day is valid if there are 10 hours of data each with at least one step), and (3) 3-a-day (eg, a day is valid if there is data within 3 periods of the day).

In the past, research-grade devices did not have the capacity to collect data over time, but with the advent of newer devices that can collect data longitudinally over several months and years, concepts of valid week or valid month have been introduced. Researchers should question, for example, how many valid days per week or month is sufficient for their specific analysis; whether valid days, weeks, or months should be consecutive and for how long; or whether valid data should be regularly occurring rather than having long-term gaps in between valid data points. All these are fitness-for-use measures unique to person-generated wearable device data, which means that depending on the research question and data type involved, the definitions for valid days, weeks, and months may differ or may not be required. The large number of potential research questions and different data types makes a one-size-fits-all approach infeasible for data completeness and suggests the need to investigate fitness-for-use measures that apply to person-generated wearable device data. Furthermore, explicitly stating the completeness definitions used in the analyses would benefit future researchers in reproducing the work. As data completeness is complex in nature, further work to assist the assessment of data completeness would alleviate the burden on researchers.

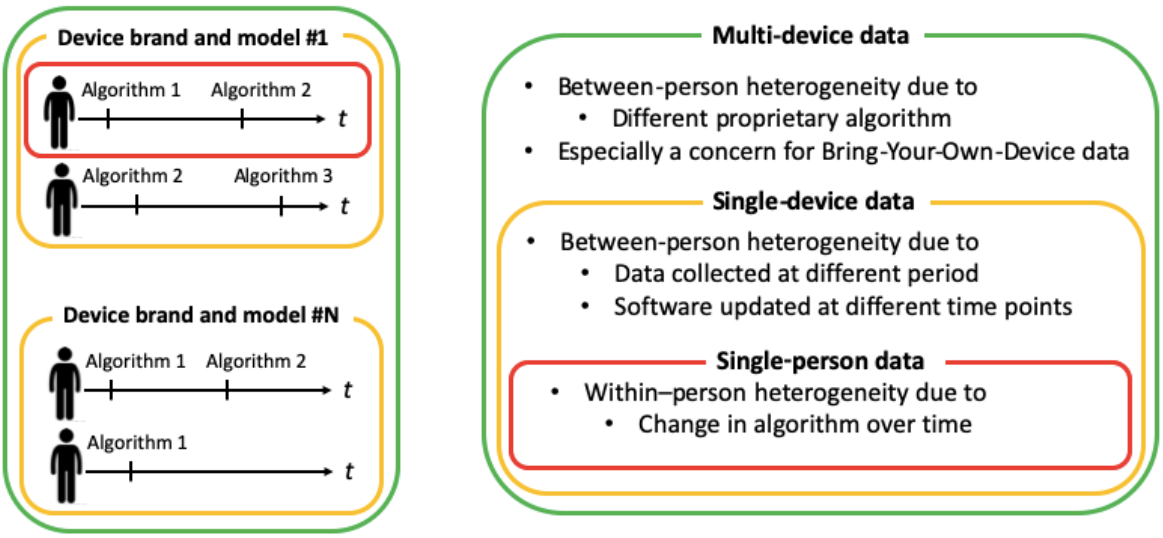
Data Correctness

Checking the correctness of data values is another quality-related challenge, as it is impossible to retrospectively identify the correct value. This is especially the case for data generated through the routine use of wearable devices because it is unlikely that a gold standard data set would exist. One approach to circumvent this challenge might be to identify outliers that are against common sense and rules for plausibility based on published values in the literature. An example rule would be that there should be no steps counted during sleep mode. The fact that researchers are currently using ad hoc rules can lead to inconsistencies and difficulty in replicating studies. Thus, a community effort to create a knowledge base for data quality rules would be beneficial to researchers because creating data quality rules is time consuming and heavily dependent on domain experts. Another indirect method to speculate data correctness would be to assess the concordance of user input data, such as age, gender, height, and weight, with another data source such as the EHR. It is known that incorrect user input while setting up the device may result in incorrect data values, as there are variables calculated based on user input (eg, calorie expenditure) [50]. If the demographic data recorded in the wearable device and the EHR agree with each other, we can at least be assured that the data values were calculated based on a trustable user input. This is an important step for those who are interested in using both wearable device data and EHR data in their study.

Data Heterogeneity

Through this review, the authors found that heterogeneity of data exists at three levels—single-person data (a data set generated by a single person), single-device data (data set generated by multiple people who use the same brand, model, and version of device; eg, a data set consisting of data generated from Fitbit Charge HR), and multidevice data (a data set generated by multiple people who use diverse brands, models, and versions of devices, eg, data set consisting of data generated from Fitbit Charge HR, Fitbit Alta HR, Withings Steel HR Sport, Apple Watch Series 3, etc). Figure 2 depicts the three levels of data heterogeneity.

Figure 2. Data heterogeneity on three levels.



In single-person data, a change in algorithms over time may produce within-person heterogeneity [24]. For single-device data, there would be between-person heterogeneity, as data are collected from multiple people at different periods, where different versions of algorithms can be used across people depending on the period of data collection [24]. Even if data

are collected in the same period, heterogeneity could exist if the software is updated at different time points across individuals. In this setting, both between-person and within-person heterogeneity can occur simultaneously. For multidevice data, the heterogeneity increases even more because of the different proprietary algorithms used for different devices. There would be between-person heterogeneity across data from individuals using different devices in addition to the between-person heterogeneity across data from individuals using the same device and within-person heterogeneity across data from different time points within the same person. This would especially be a concern for data sets collected under a bring-your-own-device policy, as individuals would provide data from different device brands, models, and different periods. Thus, it is recommended that data providers collect metadata on which device brand, model, and software version the data are generated from, and researchers should check this metadata before conducting their analyses. It would also be a good practice to report these data when publishing study results so that they could be compared with other studies [51,52].

Through the literature review process, we found that there is a lack of studies that thoroughly investigate the data quality challenges of person-generated wearable device data, especially for research purposes. Although the current literature describes the existence of data quality problems, it rarely elaborates on how the data quality metrics were defined or how the data quality problems of wearable device data were assessed. For large-scale, routinely collected wearable device data that are commonly used for biomedical research, further studies are needed to deeply understand the data quality challenges for wearable device data and provide guidance to researchers.

Limitations

One limitation of this study is that only one researcher went through the process of screening the title and abstract of studies. Therefore, the selection of literature could have been subjective in the initial phase of screening, and there is the possibility that some factors or challenges were not extracted because of potential biases in selecting the literature. However, the reviewer followed the systematic, a priori–defined selection criteria and data extraction rules to ensure consistency and reproducibility [53]. Although the initial screening of the literature was performed by a single author, other activities such as full-text screening, determining search queries, and categorizing extracted data were conducted by multiple authors. Another limitation is that although we excluded device validation studies in our review, these studies may mention factors affecting data quality for research. However, our full-text screening contained a few device validation studies, and we did not find unique information that was not captured from the final list of 19 studies.

Conclusions

The goals of this review were to (1) summarize the factors associated with data quality reported in the literature with respect to passive data collection methods using wearable devices, (2) identify data quality challenges of wearable device data, and (3) deduce implications on data quality challenges for using data for research purposes. With this goal in mind, we identified three categories—namely device- and technical-related, user-related, and data governance-related factors—along with the associated data quality problems mentioned in the literature—namely completeness, correctness, and heterogeneity. In the case of the secondary use of data, knowing the factors may not directly help researchers, as most of the problems cannot be retrospectively amended. However, the value of this study is that it facilitates the understanding of the potential causes of data quality challenges, which is a complex and time-consuming process that requires thorough discussions among domain experts, analysts, and researchers [45,54]. Moreover, it could guide the application of appropriate analytical procedures to mitigate the negative impact on analytic results. Our review provides some insight into potential data quality problems, such as the incorrectness, incompleteness, and heterogeneity of data. However, further work is required to gain a deeper understanding of each challenge, to investigate if there are any other existing challenges that have not been discovered in the literature, and to provide guidance on data quality assessments for person-generated wearable device data.

Acknowledgments

This research was supported by the NIH (grant numbers: 1U01TR002062-01 and 1U2COD023196-01).

Authors' Contributions

SC and KN designed the study. SC and KN screened, reviewed, and extracted data from the literature. SC, KN, and IE conducted multiple iterations to categorize the extracted data. SC drafted the manuscript. KN, IE, CW, and MK reviewed, edited, and provided feedback on the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms used in scholarly databases and a summary of studies included in this review.
[DOCX File , 31 KB - mhealth_v9i3e20738_app1.docx]

References

1. Nittas V, Mütsch M, Ehrler F, Puhan MA. Electronic patient-generated health data to facilitate prevention and health promotion: a scoping review protocol. BMJ Open 2018 Aug 10;8(8):e021245 [FREE Full text] [doi: 10.1136/bmjopen-2017-021245] [Medline: 30099392]
2. McCarthy J. One in Five U.S. Adults Use Health Apps, Wearable Trackers. Gallup. 2019 Dec 11. URL: https://news.gallup.com/poll/269096/one-five-adults-health-apps-wearable-trackers.aspx [accessed 2020-09-20]
3. Wood WA, Bennett AV, Basch E. Emerging uses of patient generated health data in clinical research. Mol Oncol 2015 May;9(5):1018-1024 [FREE Full text] [doi: 10.1016/j.molonc.2014.08.006] [Medline: 25248998]
4. Hickey AM, Freedson PS. Utility of Consumer Physical Activity Trackers as an Intervention Tool in Cardiovascular Disease Prevention and Treatment. Prog Cardiovasc Dis 2016;58(6):613-619. [doi: 10.1016/j.pcad.2016.02.006] [Medline: 26943981]
5. 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]
6. Izmailova ES, Wagner JA, Perakslis ED. Wearable Devices in Clinical Trials: Hype and Hypothesis. Clin Pharmacol Ther 2018 Jul;104(1):42-52 [FREE Full text] [doi: 10.1002/cpt.966] [Medline: 29205294]
7. Codella J, Partovian C, Chang HY, Chen CH. Data quality challenges for person-generated health and wellness data. IBM J. Res. & Dev 2018 Jan 1;62(1):3:1-3:8. [doi: 10.1147/jrd.2017.2762218]
8. Lim WK, Davila S, Teo JX, Yang C, Pua CJ, Blöcker C, et al. Beyond fitness tracking: The use of consumer-grade wearable data from normal volunteers in cardiovascular and lipidomics research. PLoS Biol 2018 Dec;16(2):e2004285 [FREE Full text] [doi: 10.1371/journal.pbio.2004285] [Medline: 29485983]
9. McDonald L, Mehmud F, Ramagopalan SV. Sleep and BMI: Do (Fitbit) bands aid? F1000Res 2018;7:511 [FREE Full text] [doi: 10.12688/f1000research.14774.2] [Medline: 30271578]
10. Cheung YK, Hsueh PS, Ensari I, Willey JZ, Diaz KM. Quantile Coarsening Analysis of High-Volume Wearable Activity Data in a Longitudinal Observational Study. Sensors (Basel) 2018 Sep 12;18(9) [FREE Full text] [doi: 10.3390/s18093056] [Medline: 30213093]
11. Burg MM, Schwartz JE, Kronish IM, Diaz KM, Alcantara C, Duer-Hefele J, et al. Does Stress Result in You Exercising Less? Or Does Exercising Result in You Being Less Stressed? Or Is It Both? Testing the Bi-directional Stress-Exercise Association at the Group and Person (N of 1) Level. Ann Behav Med 2017 Dec;51(6):799-809 [FREE Full text] [doi: 10.1007/s12160-017-9902-4] [Medline: 28290065]
12. Henriksen A, Haugen MM, Woldaregay AZ, Muzny M, Hartvigsen G, Hopstock LA, et al. Using Fitness Trackers and Smartwatches to Measure Physical Activity in Research: Analysis of Consumer Wrist-Worn Wearables. J Med Internet Res 2018 Mar 22;20(3):e110 [FREE Full text] [doi: 10.2196/jmir.9157] [Medline: 29567635]
13. Safran C. Update on Data Reuse in Health Care. Yearb Med Inform 2017 Aug;26(1):24-27 [FREE Full text] [doi: 10.15265/IY-2017-013] [Medline: 29063535]
14. National Institutes of Health (NIH). Data Sources. All of Us. URL: https://www.researchallofus.org/data/data-sources/#mobile-health-data [accessed 2021-02-27]
15. Menai M, Brouard B, Vegreville M, Chieh A, Schmidt N, Oppert J, et al. Cross-Sectional and longitudinal associations of objectively-measured physical activity on blood pressure: evaluation in 37 countries. Health Promot Perspect 2017;7(4):190-196 [FREE Full text] [doi: 10.15171/hpp.2017.34] [Medline: 29085795]
16. Kim K, Nikzad N, Quer G, Wineinger NE, Vegreville M, Normand A, et al. Real World Home Blood Pressure Variability in Over 56,000 Individuals With Nearly 17 Million Measurements. Am J Hypertens 2018 Apr 13;31(5):566-573 [FREE Full text] [doi: 10.1093/ajh/hpx221] [Medline: 29365036]
17. Weiner MG, Embi PJ. Toward reuse of clinical data for research and quality improvement: the end of the beginning? Ann Intern Med 2009 Sep 01;151(5):359-360. [doi: 10.7326/0003-4819-151-5-200909010-00141] [Medline: 19638404]
18. Zozus M, Kahn M, Weiskopf N. Data Quality in Clinical Research. In: Clinical Research Informatics. New York, NY: Springer International Publishing; 2019:213-248.
19. 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;2:45 [FREE Full text] [doi: 10.1038/s41746-019-0121-1] [Medline: 31304391]
20. Abdolkhani R, Borda A, Gray K. Quality Management of Patient Generated Health Data in Remote Patient Monitoring Using Medical Wearables - A Systematic Review. Stud Health Technol Inform 2018;252:1-7. [Medline: 30040674]
21. Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. EGEMS (Wash DC) 2016;4(1):1244 [FREE Full text] [doi: 10.13063/2327-9214.1244] [Medline: 27713905]
22. Fiorini N, Canese K, Starchenko G, Kireev E, Kim W, Miller V, et al. Best Match: New relevance search for PubMed. PLoS Biol 2018 Aug;16(8):e2005343 [FREE Full text] [doi: 10.1371/journal.pbio.2005343] [Medline: 30153250]
23. Piasecki J, Waligora M, Dranseika V. Google Search as an Additional Source in Systematic Reviews. Sci Eng Ethics 2018 Apr;24(2):809-810 [FREE Full text] [doi: 10.1007/s11948-017-0010-4] [Medline: 29249022]

24. Wright SP, Hall BTS, Collier SR, Sandberg K. How consumer physical activity monitors could transform human physiology research. Am J Physiol Regul Integr Comp Physiol 2017 Mar 01;312(3):R358-R367. [doi: 10.1152/ajpregu.00349.2016] [Medline: 28052867]
25. Bietz MJ, Bloss CS, Calvert S, Godino JG, Gregory J, Claffey MP, et al. Opportunities and challenges in the use of personal health data for health research. J Am Med Inform Assoc 2016 Apr;23(e1):e42-e48. [doi: 10.1093/jamia/ocv118] [Medline: 26335984]
26. Fawcett T. Mining the Quantified Self: Personal Knowledge Discovery as a Challenge for Data Science. Big Data 2015 Dec;3(4):249-266. [doi: 10.1089/big.2015.0049] [Medline: 27441406]
27. Düking P, Fuss FK, Holmberg H, Sperlich B. Recommendations for Assessment of the Reliability, Sensitivity, and Validity of Data Provided by Wearable Sensors Designed for Monitoring Physical Activity. JMIR Mhealth Uhealth 2018 Apr 30;6(4):e102 [FREE Full text] [doi: 10.2196/mhealth.9341] [Medline: 29712629]
28. Guo A, Ma J. Context-Aware Scheduling in Personal Data Collection From Multiple Wearable Devices. IEEE Access 2017;5:2602-2614. [doi: 10.1109/access.2017.2666419]
29. Constantinou V, Felber AE, Chan JL. Applicability of consumer activity monitor data in marathon events: an exploratory study. J Med Eng Technol 2017 Oct;41(7):534-540. [doi: 10.1080/03091902.2017.1366560] [Medline: 28954563]
30. Hardy J, Veinot TC, Yan X, Berrocal VJ, Clarke P, Goodspeed R, et al. User acceptance of location-tracking technologies in health research: Implications for study design and data quality. J Biomed Inform 2018 Mar;79:7-19 [FREE Full text] [doi: 10.1016/j.jbi.2018.01.003] [Medline: 29355784]
31. Banerjee T, Sheth A. IoT Quality Control for Data and Application Needs. IEEE Intell. Syst 2017 Mar;32(2):68-73. [doi: 10.1109/mis.2017.35]
32. Karkouch A, Mousannif H, Al Moatassime H, Noel T. Data quality in internet of things: A state-of-the-art survey. Journal of Network and Computer Applications 2016 Sep;73:57-81. [doi: 10.1016/j.jnca.2016.08.002]
33. Liang Z, Ploderer B, Chapa-Martell M. Is Fitbit Fit for Sleep-tracking? Sources of Measurement Errors and Proposed Countermeasures. USA: ACM; 2017 Presented at: Proc 11th EAI Int Conf Pervasive Comput Technol Healthc; May 23-26, 2017; New York, NY p. 476-479. [doi: 10.1145/3154862.3154897]
34. Reinerman-Jones L, Harris J, Watson A. Considerations for using fitness trackers in psychophysiology research. In: Human Interface and the Management of Information: Information, Knowledge and Interaction Design. HIMI 2017. New York, NY: Springer; 2017:598-606.
35. Jülicher T, Delisle M. Step into "The Circle" - A Close Look at Wearables and Quantified Self. In: Big Data Context. New York, NY: Springer; 2018:81-91.
36. Beukenhorst AL, Sergeant JC, Little MA, McBeth J, Dixon WG. Consumer Smartwatches for Collecting Self-Report and Sensor Data: App Design and Engagement. Stud Health Technol Inform 2018;247:291-295. [Medline: 29677969]
37. Cleland I, Donnelly MP, Nugent CD, Hallberg J, Espinilla M, Garcia-Constantino M. Collection of a Diverse, Realistic and Annotated Dataset for Wearable Activity Recognition. 2018 Presented at: 2018 IEEE International Conference on Pervasive Computing and Communications Workshops; March 19-23, 2018; Athens, Greece p. 555-560.
38. Mezghani E, Exposito E, Drira K, Da Silveira M, Pruski C. A Semantic Big Data Platform for Integrating Heterogeneous Wearable Data in Healthcare. J Med Syst 2015 Dec;39(12):185. [doi: 10.1007/s10916-015-0344-x] [Medline: 26490143]
39. Lai AM, Hsueh PS, Choi YK, Austin RR. Present and Future Trends in Consumer Health Informatics and Patient-Generated Health Data. Yearb Med Inform 2017 Aug;26(1):152-159 [FREE Full text] [doi: 10.15265/IY-2017-016] [Medline: 29063559]
40. Oh J, Lee U. Exploring UX issues in Quantified Self technologies. 2015 Presented at: 2015 Eighth International Conference on Mobile Computing and Ubiquitous Networking (ICMU); January 20-22, 2015; Hakodate, Japan p. 53-59. [doi: 10.1109/icmu.2015.7061028]
41. Wickel EE. Reporting the reliability of accelerometer data with and without missing values. PLoS One 2014;9(12):e114402 [FREE Full text] [doi: 10.1371/journal.pone.0114402] [Medline: 25478692]
42. Stephens S, Beyene J, Tremblay MS, Faulkner G, Pullnayegum E, Feldman BM. Strategies for Dealing with Missing Accelerometer Data. Rheum Dis Clin North Am 2018 May;44(2):317-326. [doi: 10.1016/j.rdc.2018.01.012] [Medline: 29622298]
43. Sun H. Enterprise Information Management: Best Practices in Data Governance.: Oracle Corporation; 2011 May. URL: https://www.oracle.com/technetwork/articles/entarch/oea-best-practices-data-gov-400760.pdf
44. West P, Van KM, 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]
45. Khare R, Ruth B, Miller M, Tucker J, Utidjian L, Razzaghi H. Predicting Causes of Data Quality Issues in a Clinical Data Research Network. 2018 May 18 Presented at: AMIA Joint Summits Translational Science; March 12-15, 2018; San Francisco p. 113-121.
46. Ong T, Pradhananga R, Holve E, Kahn MG. A Framework for Classification of Electronic Health Data Extraction-Transformation-Loading Challenges in Data Network Participation. EGEMS (Wash DC) 2017 Jun 13;5(1):10 [FREE Full text] [doi: 10.5334/egems.222] [Medline: 29930958]

47. Mailey EL, Gothe NP, Wójcicki TR, Szabo AN, Olson EA, Mullen SP, et al. Influence of allowable interruption period on estimates of accelerometer wear time and sedentary time in older adults. J Aging Phys Act 2014 Apr;22(2):255-260 [FREE Full text] [doi: 10.1123/japa.2013-0021] [Medline: 23752299]
48. Evenson KR, Terry JW. Assessment of differing definitions of accelerometer nonwear time. Res Q Exerc Sport 2009 Jun;80(2):355-362 [FREE Full text] [doi: 10.1080/02701367.2009.10599570] [Medline: 19650401]
49. Tang LM, Meyer J, Epstein DA, Bragg K, Engelen L, Bauman A, et al. Defining Adherence: Making Sense of Physical Activity Tracker Data. ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2018 Mar;2(1):1-22. [doi: 10.1145/3191769]
50. Low CA, Dey AK, Ferreira D, Kamarck T, Sun W, Bae S, et al. Estimation of Symptom Severity During Chemotherapy From Passively Sensed Data: Exploratory Study. J Med Internet Res 2017;19(12):e420 [FREE Full text] [doi: 10.2196/jmir.9046] [Medline: 29258977]
51. Collins T, Woolley SI, Oniani S, Pires IM, Garcia NM, Ledger SJ, et al. Version Reporting and Assessment Approaches for New and Updated Activity and Heart Rate Monitors. Sensors (Basel) 2019 Apr 10;19(7) [FREE Full text] [doi: 10.3390/s19071705] [Medline: 30974755]
52. 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(1):159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
53. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009;339:b2700 [FREE Full text] [Medline: 19622552]
54. Ancker JS, Shih S, Singh MP, Snyder A, Edwards A, Kaushal R, HITEC investigators. Root causes underlying challenges to secondary use of data. AMIA Annu Symp Proc 2011;2011:57-62 [FREE Full text] [Medline: 22195055]

Abbreviations

ACM: Association for Computing Machinery
EHR: electronic health record
ETL: extract, transform, and load
IEEE: Institute of Electrical and Electronics Engineers
NIH: National Institutes of Health
PGHD: person-generated health data

Edited by L Buis; submitted 04.07.20; peer-reviewed by R Walmsley, B Hanson ; comments to author 14.09.20; revised version received 07.12.20; accepted 18.02.21; published 19.03.21.

Please cite as:
Cho S, Ensari I, Weng C, Kahn MG, Natarajan K
Factors Affecting the Quality of Person-Generated Wearable Device Data and Associated Challenges: Rapid Systematic Review
JMIR Mhealth Uhealth 2021;9(3):e20738
URL: https://mhealth.jmir.org/2021/3/e20738
doi:10.2196/20738
PMID:33739294

Original Paper

Measurement of Human Walking Movements by Using a Mobile Health App: Motion Sensor Data Analysis

Sungchul Lee[1], PhD; Ryan M Walker[2], MSc; Yoohwan Kim[2], PhD; Hyunhwa Lee[3], MSN, APRN, PhD

[1]School of Computing and Information Systems, Grand Valley State University, Allendale, MI, United States

[2]Computer Science, University of Nevada, Las Vegas, Las Vegas, NV, United States

[3]School of Nursing, University of Nevada, Las Vegas, Las Vegas, NV, United States

Corresponding Author:
Sungchul Lee, PhD
School of Computing and Information Systems
Grand Valley State University
D-2-206 Mackinac Hall
Allendale, MI, 49401
United States
Phone: 1 616 331 4372
Email: lees2@gvsu.edu

Abstract

Background: This study presents a new approach to measure and analyze the walking balance of humans by collecting motion sensor data in a smartphone.

Objective: We aimed to develop a mobile health (mHealth) app that can measure the walking movements of human individuals and analyze the differences in the walking movements of different individuals based on their health conditions. A smartphone's motion sensors were used to measure the walking movements and analyze the rotation matrix data by calculating the variation of each xyz rotation, which shows the variables in walking-related movement data over time.

Methods: Data were collected from 3 participants, that is, 2 healthy individuals (1 female and 1 male) and 1 male with back pain. The participant with back pain injured his back during strenuous exercise but he did not have any issues in walking. The participants wore the smartphone in the middle of their waistline (as the center of gravity) while walking. They were instructed to walk straight at their own pace in an indoor hallway of a building. The walked a distance of approximately 400 feet. They walked for 2-3 minutes in a straight line and then returned to the starting location. A rotation vector in the smartphone, calculated by the rotation matrix, was used to measure the pitch, roll, and yaw angles of the human body while walking. Each xyz-rotation vector datum was recalculated to find the variation in each participant's walking movement.

Results: The male participant with back pain showed a diminished level of walking balance with a wider range of xyz-axis variations in the rotations compared to those of the healthy participants. The standard deviation in the xyz-axis of the male participant with back pain was larger than that of the healthy male participant. Moreover, the participant with back pain had the widest combined range of right-to-left and forward-to-backward motions. The healthy male participant showed smaller standard deviation while walking than the male participant with back pain and the female healthy participant, indicating that the healthy male participant had a well-balanced walking movement. The walking movement of the female healthy participant showed symmetry in the left-to-right (x-axis) and up-to-down (y-axis) motions in the x-y variations of rotation vectors, indicating that she had lesser bias in gait than the others.

Conclusions: This study shows that our mHealth app based on smartphone sensors and rotation vectors can measure the variations in the walking movements of different individuals. Further studies are needed to measure and compare walking movements by age, gender, as well as types of health problems or disease. This app can help in finding differences in gait in people with diseases that affect gait.

(JMIR Mhealth Uhealth 2021;9(3):e24194) doi:10.2196/24194

KEYWORDS

mobile health; mHealth; walking balance; smartphone; motion sensor; sensor; walking; walking balance; mobile phone

Introduction

Human balance is achieved and maintained by a complex set of human sensorimotor and musculoskeletal systems that control vision, proprioception, vestibular function, muscle contraction, and others. Multiple factors such as psychological factors, injury, or disease can affect these components [1]. Postural balance is commonly used for measurement in healthy and pathological participants and is used for diagnosing disorders related to the nervous system [2], such as ataxia [3], cognitive deficits [4-6], Parkinson disease [7,8], vision problems [9], Alzheimer [4], and so on. Walking balance is also a good method for measuring human postural balance, as it requires the coordinated use of the visual, vestibular, and musculoskeletal systems [10]. Walking balance can become less stable and can fluctuate if an individual has experienced a stroke [11] or a lower limb or back injury [12] due to fragile biomechanical structures in the sensorimotor and musculoskeletal systems that influence how the human body moves while walking [13,14].

Mobile health (mHealth) supports methods that measure physical activities [15] such as measuring human balance and stability by using gravity, linear acceleration, and orientation [16,17]. Moreover, in previous mHealth research, smartphones have been used to support the diagnosis of diseases related to human balance, such as Parkinson disease [18]. However, those studies did not fully utilize smartphone sensors even though smartphones have multiple physical and software sensors. Those studies have also not considered the previous steps even though human steps affect the next step. Additionally, the x, y, and z axes of the motion sensors should be analyzed to obtain a detailed understanding of the human walking movement.

Our study introduces a new method to measure the walking balance by using motion sensors in a smartphone to determine the rotation vector. The rotation vector provides the pitch, roll, and yaw angles in the smartphone. We can, therefore, measure the pitch, roll, and yaw angles of the human body while walking by using the smartphone worn around the body waistline. Data were collected from 1 healthy female, 1 healthy male, and 1 male with back pain wearing the smartphone while walking. The male experiencing back pain injured his back during exercise but he did not have any issues walking. His back pain was confirmed by his doctor's note.

We developed an mHealth app to record and analyze the sensor output in the smartphone while participants walk. Pitch, roll, and yaw angles were extrapolated from the recorded sensor data, and a model was created to compare the differences in the walking balance among the participants. We used the ggplot2 graphics package in R programming language to create the visualizations.

Methods

Implementation of the mHealth App

The mHealth app was developed to measure and record rotational data in real time by using an Android smartphone's motion sensors. Figures 1-7 show the mHealth app in use to gather the sensor data. The app was programmed for Android mobile platforms with software development kits greater than 21 using Android Studio [19]. This research used the app on the default settings of the Samsung Galaxy S8 [20] with Android 7.0 mobile operating system. The code is written in Java using the Android application programming interface [21]. This app uses the Android sensor framework to access sensor data as part of the hardware package that consists of 3 classes and 1 interface. The classes are SensorManager, Sensor, and SensorEvent. SensorManager accesses the device's sensors, Sensor obtains the list of available sensors, and SensorEvent creates the sensor object that includes the raw sensor data. The interface, SensorEventListener, receives notifications when a sensor value or accuracy changes [22].

Sensors used in this app include the gyroscope, accelerometer, gravity, and magnetic sensors. The gravity and magnetic sensors are used to calculate the rotation matrix using the getRotationMatrix method, which belongs to the SensorManager class [23]. Raw sensor data visualizations are created using the Androidplot library [24]. The collected sensor data are stored in an SQLite database [25]. Two comma-separated value files are created from the saved participant information and sensor values. These files are saved on the device for further data analysis. Figure 1 shows the main fragment in the mHealth app. In the main fragment, the mHealth obtains the participant's status such as any pain (back, leg, head, etc), any medication in the last 3 days, any problems walking, concussion experience, gender, race/ethnicity, height, and weight. After touching the "SUBMIT" button in the main fragment, the mHealth app proceeds to the fragment shown in Figure 2. After selecting the "START" button (Figure 2), mHealth changes the fragment to Stop Fragment, as shown in Figure 3. Then, the participant walks straight forward for 2-3 minutes and then returns to the starting location. After the participant returns to the starting location, the participant touches the "STOP" button, as shown in Figure 3. The walking data are recorded between the time of selecting the "START" button (Figure 2) and the time of selecting the "STOP" button (Figure 3). After touching the "STOP" button, an administrator opens the widget in the mHealth app (Figure 4) to save the data by swiping from left to right on the smartphone screen. The "Admin" button (Figure 4) is used to save the recorded walking data. When selecting the "Admin" option (Figure 4), the mHealth app moves to the Save Fragment. There is 1 checkbox (Figure 5) to confirm whether the walking data are valid or not. The checkbox is for indicating whether the test was invalidated by an interruption or an unexpected event during recording. By selecting the "SAVE" button (Figure 4), the data are saved to an SQLite database and a comma-separated value file on the local disk. The mHealth app has several functions to analyze the raw data in real time. By selecting functions such as "Metrics," "Accelerometer," and "Gyroscope" (Figure 4), the mHealth app displays the current status of the raw data, as shown in Figure 6 and Figure 7. Figure 6 shows the current accelerometer, gyroscope, and rotation matrix in real time. Figure 7 shows the graph of the sensor data.

Figure 1. Main fragment of the mobile health app.

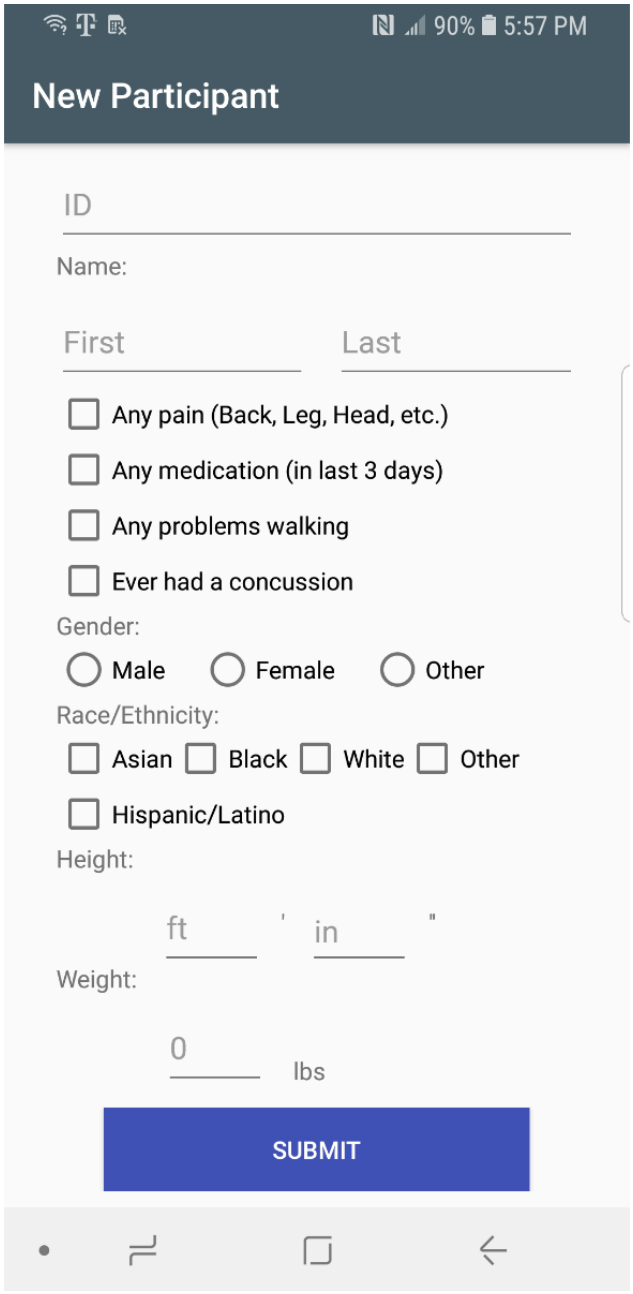


Figure 2. Start fragment of the mobile health app.

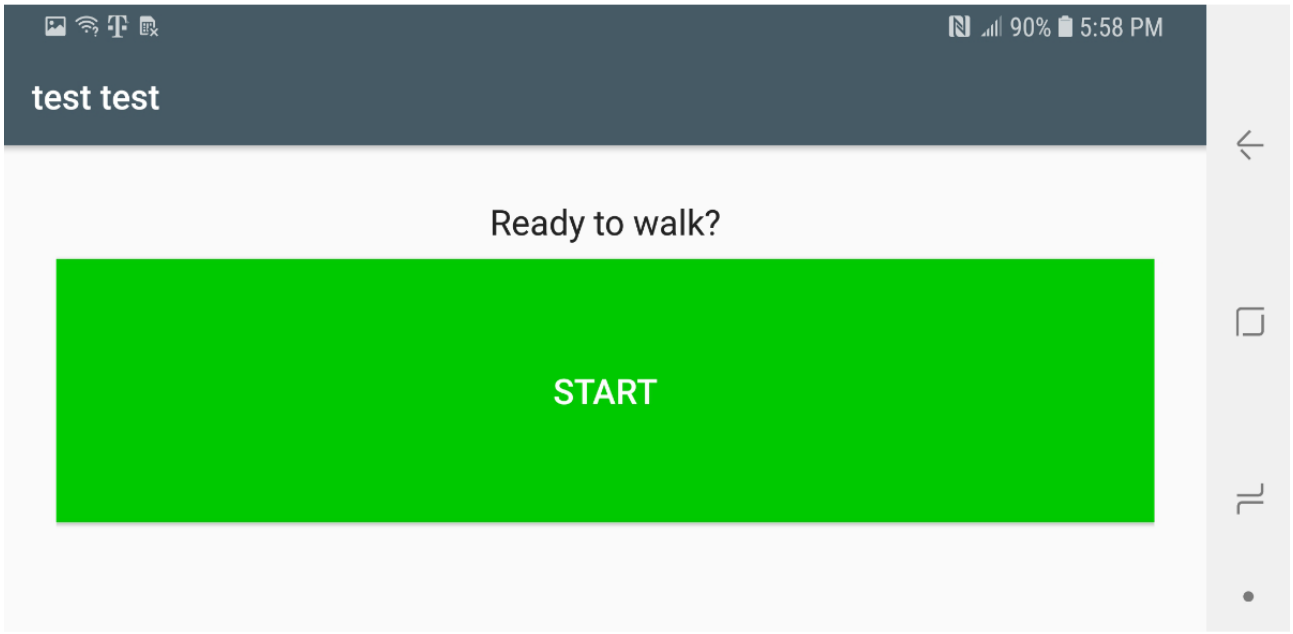


Figure 3. Stop fragment of the mobile health app.

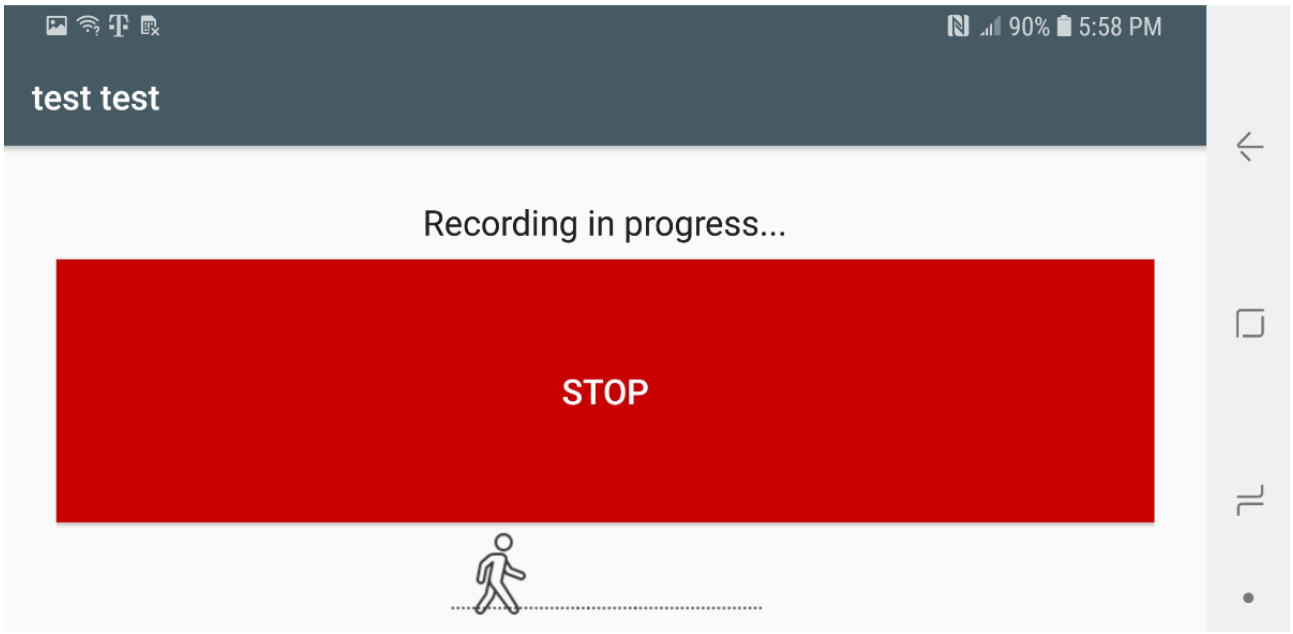


Figure 4. Menu widget.

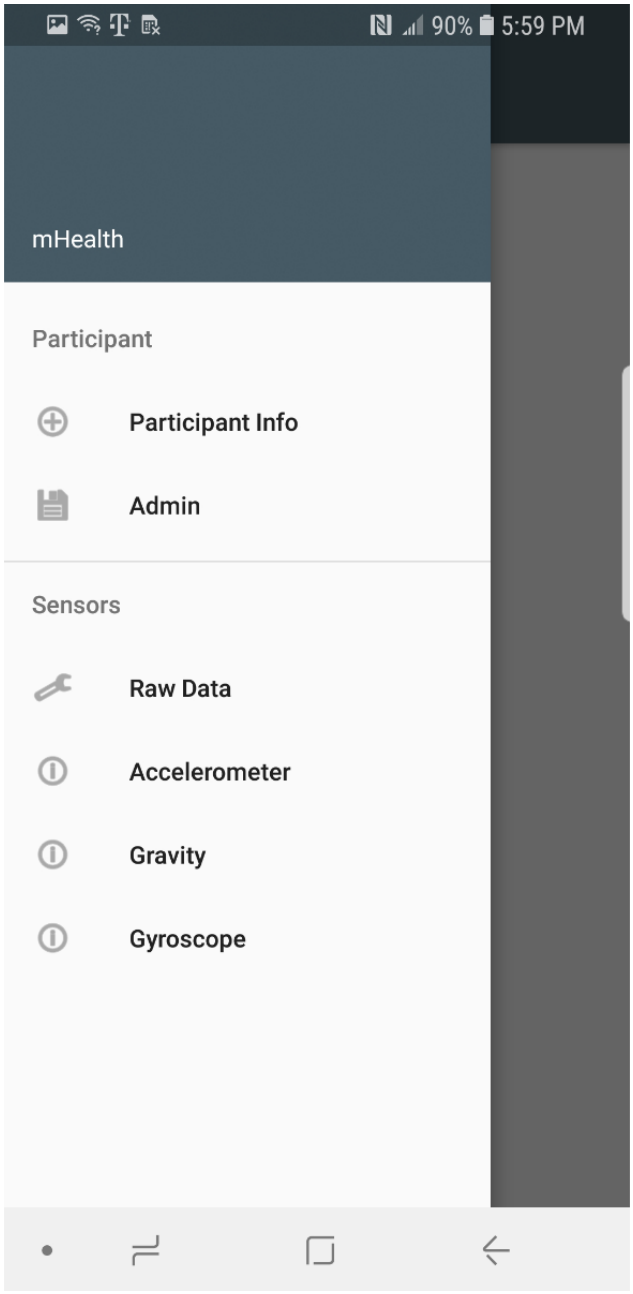


Figure 5. Confirm walk test.

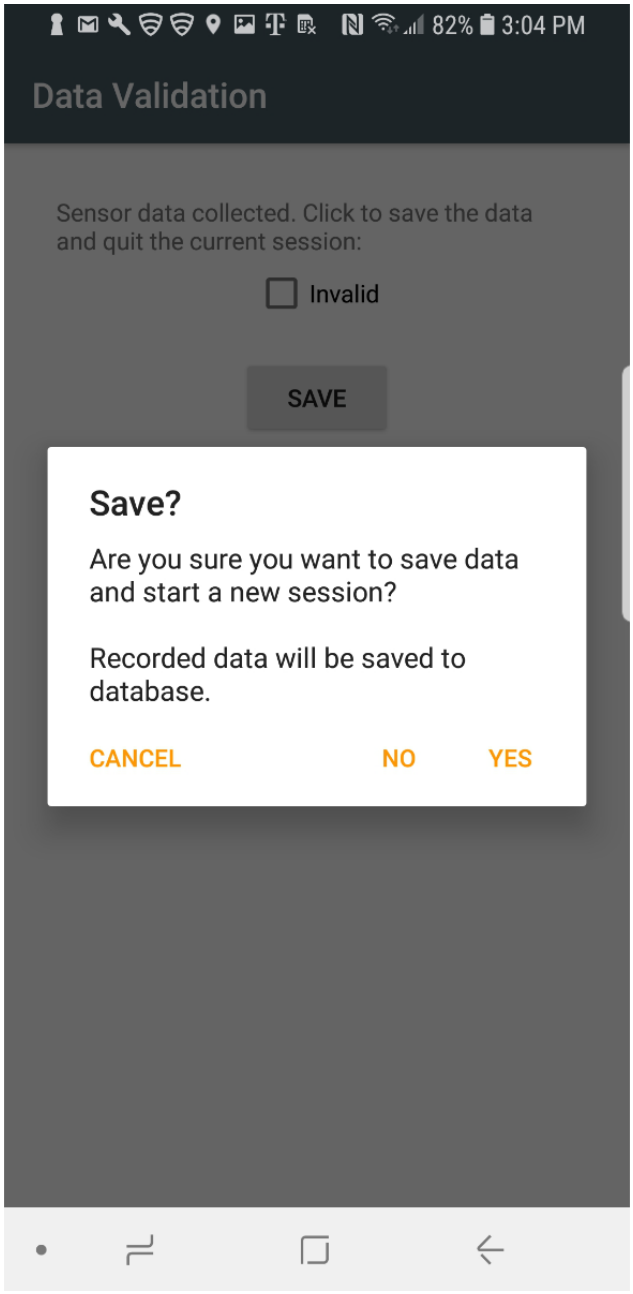


Figure 6. Sensor data matrix.

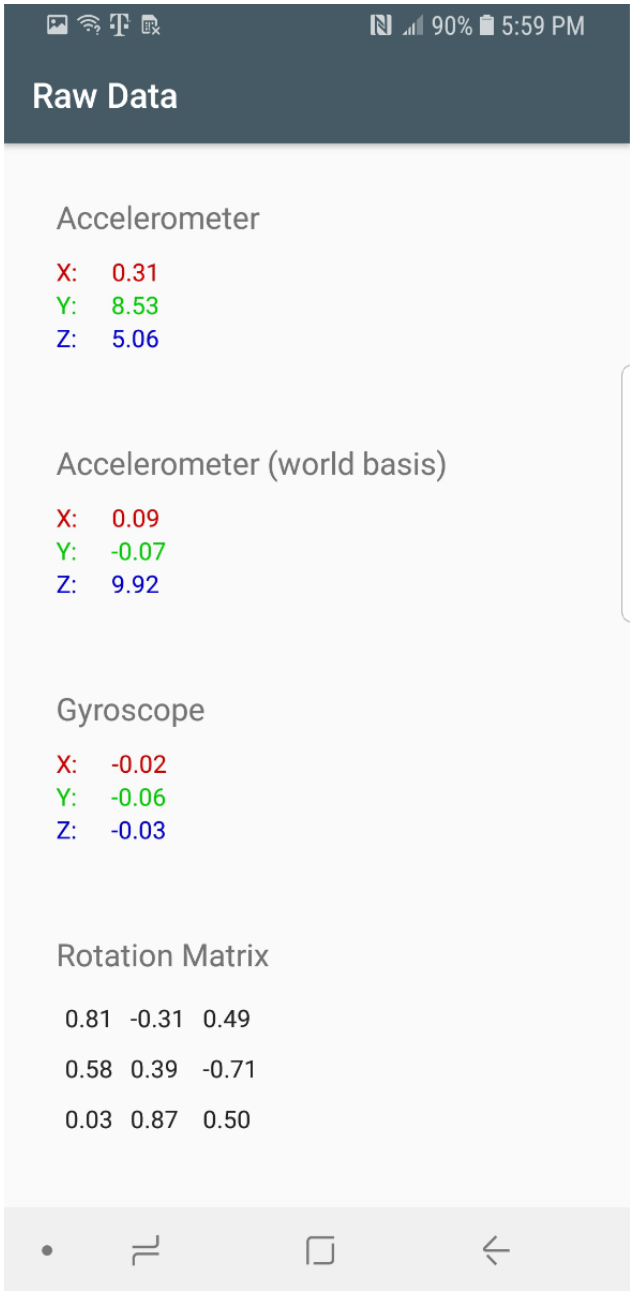
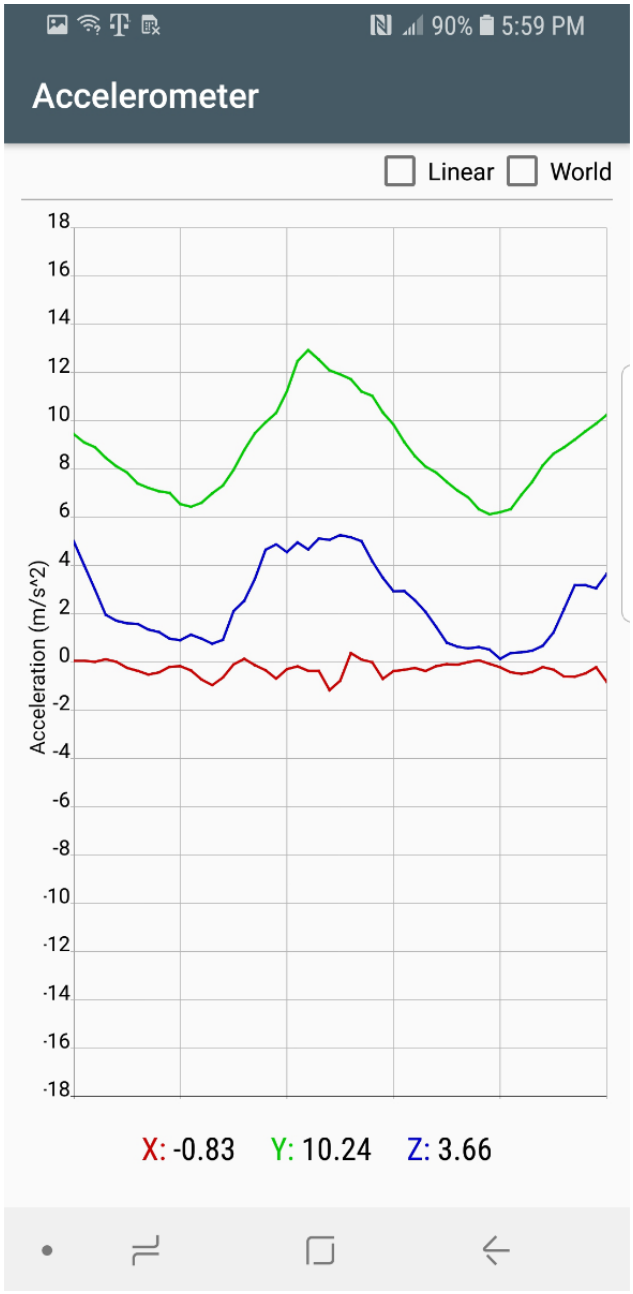


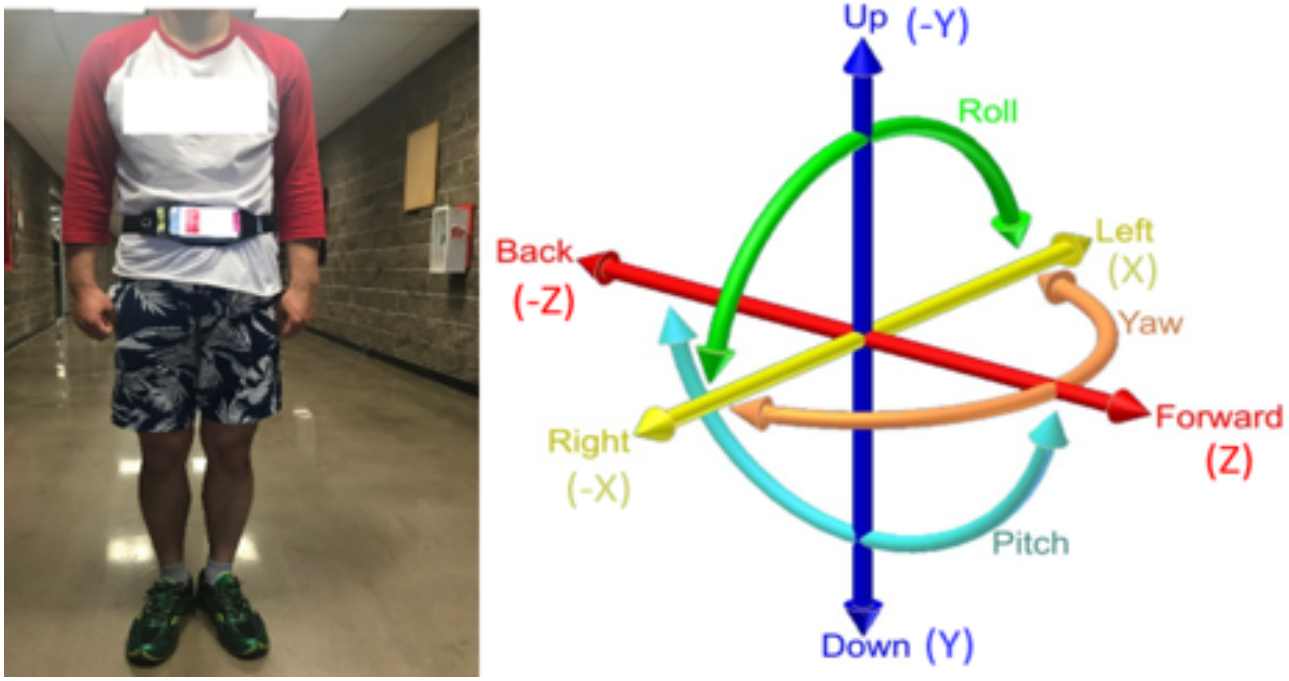
Figure 7. Sensor data graph.



Data Collection

The mHealth app was tested with 3 participants. Two participants were healthy (1 female and 1 male) and they did not have any problems walking; these 2 participants were in their twenties. The third participant with back pain was in his thirties and he had little pain in his back; however, he had no problem walking and we could not find any difference in his walking compared to the other participants. During the experiment, all the participants wore the smartphone in a pocket of a waistband at the center of their body, as shown in Figure 8. The smartphone was placed on the waistline with its top frontside facing the right side of the body. The participants were instructed to walk straight at their own pace in an indoor hallway of a building. They walked a distance of about 400 feet. They walked for about 2-3 minutes in a straight line and then returned to the starting location. The duration of the test time for each participant was 4-5 minutes. They were allowed only a single trial for the walking movement test.

Figure 8. XYZ-axis orientations of the smartphone during the walking balance test. X-axis is the right (−) and left (+) motions of the participant. Y-axis is the up (−) and down (+) motions of the participant. Z-axis is the forward (+) and backward (−) motions of the participant.



The smartphone stored the motion sensor data in the real-time mHealth app on the phone. The app collected the sensor data in 20 milliseconds. The sensor data such as accelerometer (x, y, and z), gyroscope (x, y, and z), and rotation matrix (3×3) were stored in an SQLite database and a comma-separated value file. Data corresponding to the first 10 seconds and the last 10 seconds of each participant session were removed to account for the press time of the start and stop button activation within the mHealth app. Among the sensor data, the rotation matrix was used to determine the walking balance. Because the rotation matrices can represent the rotation of the origin frame into the reference frame, the rotation matrix is commonly used to measure posture balance [26-31]. The next section explains the analysis method using the rotation matrix.

Data Analysis

Considering the differentiation in the balance control of the participants while walking, the rotation vector data are the most effective [32,33]. Therefore, the rotation vector was extrapolated from the rotation matrix data recorded with the smartphone. In the mHealth app, the xyz-axis rotation vectors of the middle of the waistline were identified as the center of gravity of each participant. The x-axis represents the body motion angle between the right and left side movements of a participant. The y-axis represents the body motion angle between the up and down movements of a participant. The z-axis represents the body motion angle between the forward and backward movements of a participant (Figure 8). The xyz-axis rotation vector was obtained from the rotation matrix. A rotation matrix is a matrix that is used to perform a rotation in Euclidean space. The Euler's theorem on the axis of a three-dimensional rotation is formulated as follows:

If R is a 3×3 orthogonal matrix ($R^TR = RR^T = I$) and R is proper (det(R) = +1), then there is a nonzero vector "v" satisfying Rv = v [34]. The rotation matrix data were collected using the rotation sensor in the Android Open Source Project. Using the rotation sensor, the mHealth app determined the rotation matrix [34] like equation 1. The rotation matrix was as follows:

From the equation 2 rotation matrix (3×3), we extracted the rotation vector x-axis (R_X), y-axis (R_Y), and z-axis (R_Z) by using the following formula [35]:

Although each participant wore the smartphone at the same location of his/her body, the sensors in the smartphone appeared to be located with slightly different slopes. Therefore, we used the difference of rotation between current time (t) and previous time (t-1) for the data analysis:

For comparison across the participants, the difference in the rotation vector values was plotted as 2D graphs using ggplot2 in R (Figures 9-11).

Figure 9. X-Y rotation vector. F1: healthy female; M1: healthy male; M2: male with back pain.

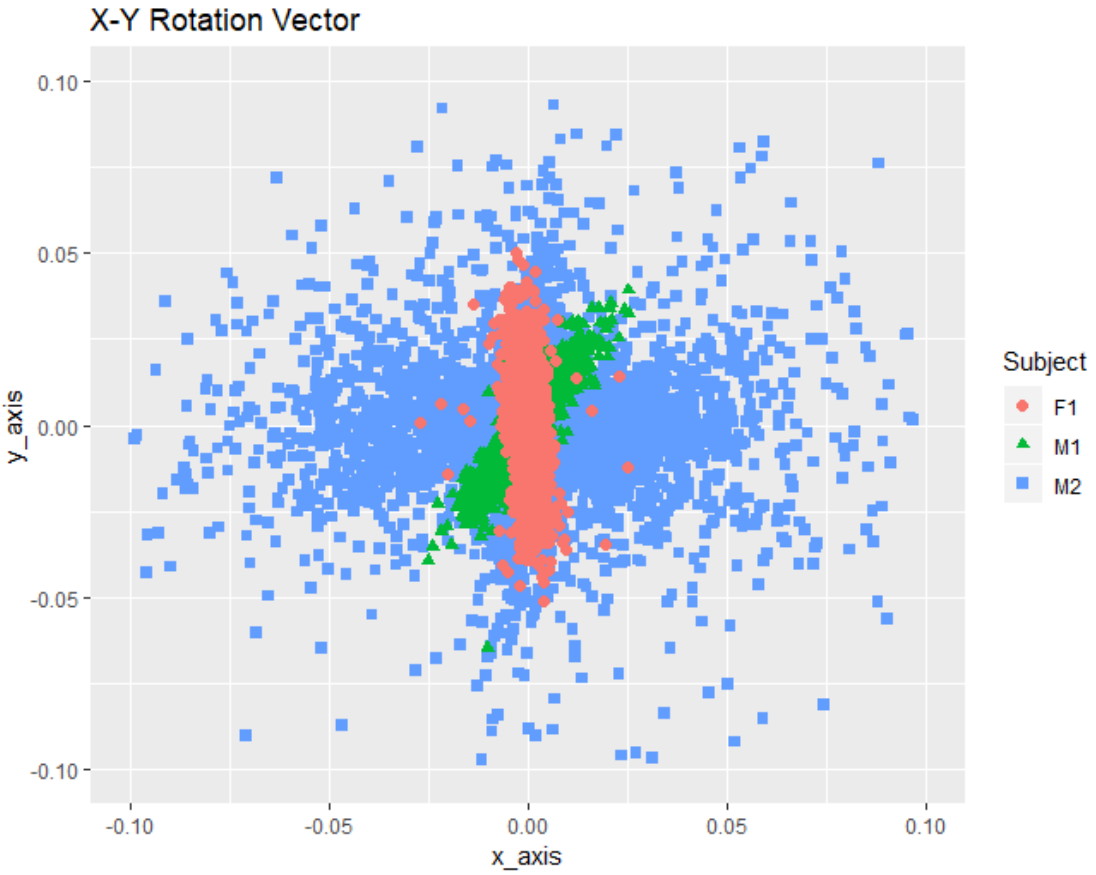


Figure 10. X-Z rotation vector. F1: healthy female; M1: healthy male; M2: male with back pain.

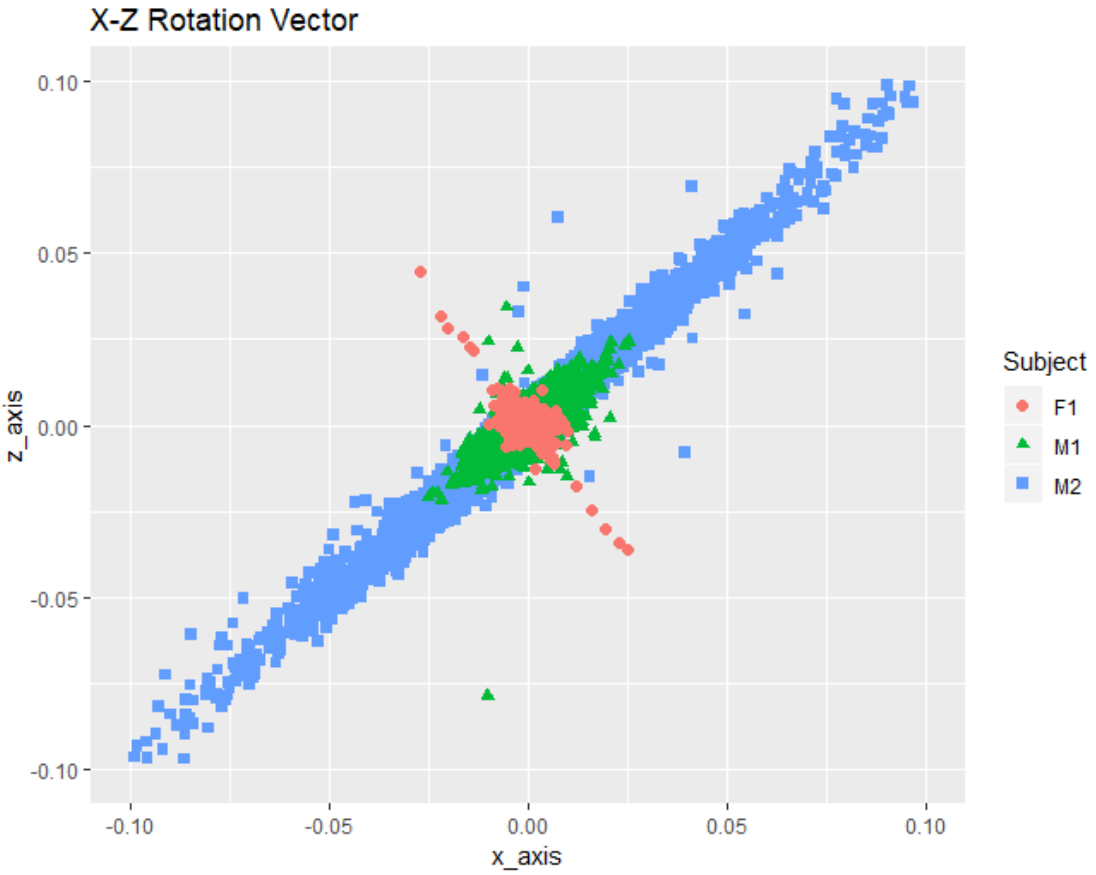
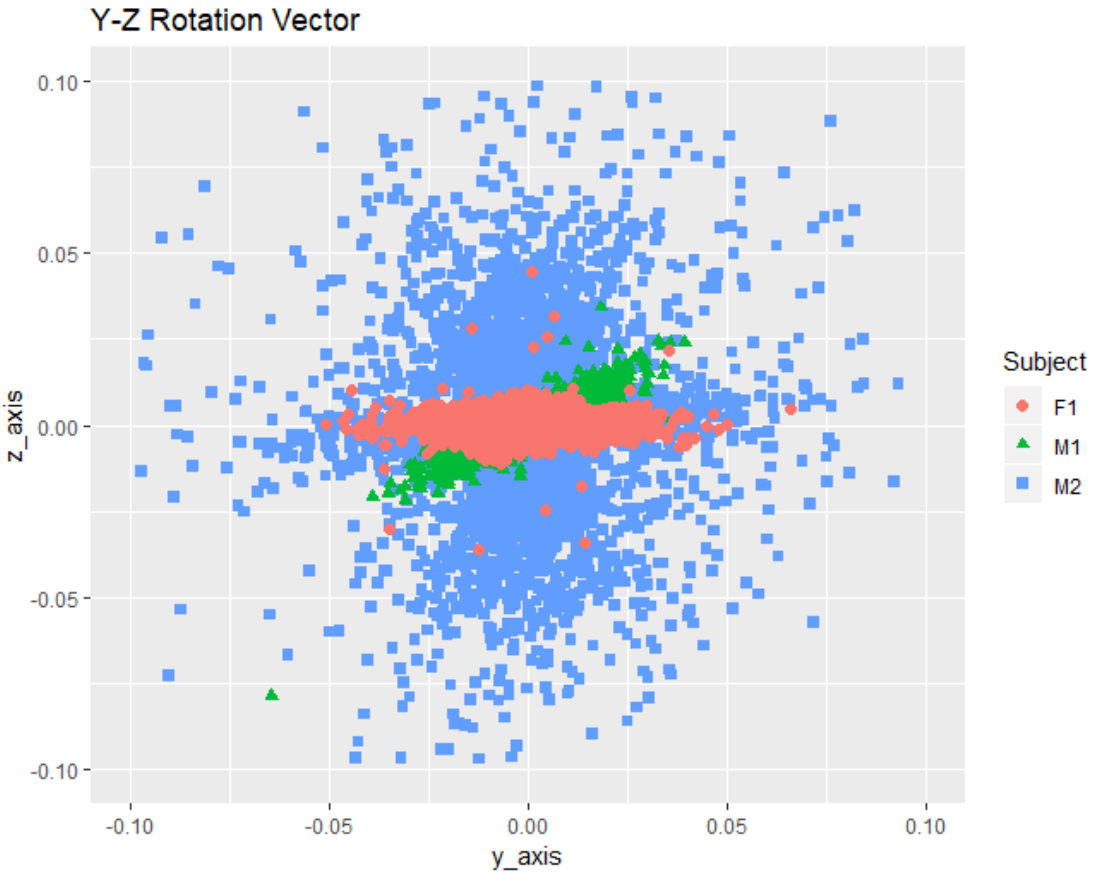


Figure 11. Y-Z rotation vector. F1: healthy female; M1: healthy male; M2: male with back pain.



Results

We collected constant motion data by using the sensors in the smartphone. Thereafter, to understand constant motion changes in the middle of the participants' waistline (umbilicus) as the center of gravity, the rotation vector values of the 3 participants were extracted and compared. There were no differences in the rotation matrix data of each participant. However, we found a difference among the participants when applying the formula (equation 4) into the data.

We calculated the variation in each step of the participants by using the formula (equation 4) (Table 1). Figures 9-11 show the illustrative rotation data, which were obtained using the formula (equation 4) by 2 combined axes during the same time period recorded by the smartphones from 2 healthy individuals and 1 individual with back pain. Illustrative rotation x-axis data show the right (−) to left (+) body motions on the x-axis. Illustrative rotation y-axis data show the up (−) to down (+) body motions on the y-axis. Illustrative rotation z-axis data recorded the forward (+) to backward (−) body motions on the z-axis. Healthy participants appeared to have narrower ranges of right-to-left and up-and-down motions than the participant with back pain who had the widest range of motion while walking. The results showed that there were differences in the body xyz-axis variations of rotation vector data between the participant with back pain and the healthy participants. The variation in the y-axis rotation of the participant with back pain was slightly wider than that of the healthy participants. Moreover, the participant with back pain appeared to have more variations in xz-axis rotations than the healthy participants, suggesting that he had a wider range of motion between right and left and forward and backward movements while walking compared to the healthy participants (Figures 9-11).

Table 1. Statistical information of the variations in the rotation vector data.

Participant, rotation vector	Maximum	Minimum	Mean	Variance	SD
Healthy male participant					
x-axis	0.025	–0.025	0	0	0.004
y-axis	0.039	–0.065	0	0	0.006
z-axis	0.034	–0.079	0	0	0.004
Healthy female participant					
x-axis	1.715	–1.715	0	0.085	0.292
y-axis	0.202	–0.21	0	0	0.017
z-axis	1.184	–1.489	0	0.036	0.188
Male participant with back pain					
x-axis	1.467	–1.466	0	0.017	0.129
y-axis	0.944	–0.534	0	0	0.021
z-axis	1.492	–1.494	0	0.017	0.13

Discussion

Principal Results

In this study, we developed an mHealth app to measure the body movements of 3 participants wearing the smartphone on their waists while walking. The motion sensors of the smartphone were used to measure the walking movements and to analyze the rotation matrix data with the proposed method, which shows the variables of walking over time. The difference in each walking step of the 3 participants was compared using a formula (equation 4). Table 1 shows the statistical information of the walking movement data after applying the formula. The healthy male participant had the smallest standard deviation (Table 1), indicating that he had the most balanced walking movement. The healthy female participant showed symmetrical walking movement in the left-to-right (x-axis) and up-and-down (y-axis) movements in the variations of the x-y axis rotation vector (Figure 9), indicating that she had lesser bias in gait than the other two participants. The x-axis standard deviation in the healthy female participant was the highest because female hip is bigger than male hip [36]. The standard deviation of y-axis and z-axis of the participant with back pain was the highest (Table 1), indicating that he had a wider movement than the other two healthy participants. Thus, the participant with back pain had a lesser balanced walking movement while walking than the other participants. Our app showed that the participant with little back pain has a wide range of movement during walking. Specifically, this participant appeared to show more variations in the right-to-left and forward-to-backward movements.

Limitations

This study has several limitations. The mHealth app was tested only in 3 participants. Nevertheless, the proposed mHealth app can effectively capture differences in the postural control during walking between healthy individuals and individuals with back pain. Further, our mHealth app is not available currently for observing the data in real time while walking. Therefore, we will implement a program that will make it possible to observe the data while walking.

Future Directions

In future research, we will conduct feasibility and efficacy testing with large pools of individuals with reduced physical mobility. For the tests, we are planning to collect walking data from high school students because we believe that this younger group is ideal for researching on walking data. This younger sample is lesser affected by aging/diseases related to walking than an older group. Further, high school students will be in school for 2 or 3 years before graduation, which is a long period for tracking their walking movements. After we collect enough walking movement data, we will compare our findings with the results from previous mHealth studies. In addition, any variable that may affect postural balance during walking such as body habitus (ie, slim vs obese) or level of daily physical activities will be considered in data collection and analysis with various sensors in smartphones. We will also consider evaluating the impact of the app at the health system level by using outcomes such as health care utilization and medication use. This study is beneficial since it provides a useful method for medical evaluation in rehabilitation and physical fitness and a means for participants to maintain a state of well-being. This research can be used to classify walking movements between people who have walking-related diseases and normal people. This classification can help in diagnosing their diseases. Thereafter, we will research the classification of walking movements based on diseases that affect walking.

Conclusions

Our study shows that mHealth and the walking rotation vector can be used to define the body walking movements. Currently, the most widely used assessments for measuring postural control are laboratory-based such as the previous walking movement research [12,13] and human balance research [2,3]. Along with requiring specialized equipment, such assessments typically do not provide real-time feedback. However, our proposed mHealth app and analysis methods support home-based measurements. The findings of this study support our mHealth app as a low-cost

and easy-to-use alternative with minimal equipment required that provides sensitive walking balance assessment [6].

Acknowledgments

This research was sponsored by the University of Nevada, Las Vegas Sports Research and Innovation Initiative Catalyst Grant under the title "Developing a low-cost mobile TBI analysis system for sports safety".

Conflicts of Interest

None declared.

References

1. Watson M, Black F, Crowson M. The human balance system: A complex coordination of central and peripheral systems. URL: https://vestibular.org/wp-content/uploads/2011/12/Human-Balance-System_36.pdf [accessed 2020-09-11]
2. Paillard T, Noé F. Techniques and Methods for Testing the Postural Function in Healthy and Pathological Subjects. Biomed Res Int 2015;2015:891390 [FREE Full text] [doi: 10.1155/2015/891390] [Medline: 26640800]
3. Bruttini C, Esposti R, Bolzoni F, Vanotti A, Mariotti C, Cavallari P. Temporal disruption of upper-limb anticipatory postural adjustments in cerebellar ataxic patients. Exp Brain Res 2015 Jan;233(1):197-203. [doi: 10.1007/s00221-014-4103-x] [Medline: 25245658]
4. Tangen G, Engedal K, Bergland A, Moger T, Mengshoel A. Relationships between balance and cognition in patients with subjective cognitive impairment, mild cognitive impairment, and Alzheimer disease. Phys Ther 2014 Aug;94(8):1123-1134. [doi: 10.2522/ptj.20130298] [Medline: 24764071]
5. Quinzaños-Fresnedo J, Fratini-Escobar PC, Almaguer-Benavides KM, Aguirre-Güemez AV, Barrera-Ortíz A, Pérez-Zavala R, et al. Prognostic validity of a clinical trunk control test for independence and walking in individuals with spinal cord injury. J Spinal Cord Med 2020 May;43(3):331-338 [FREE Full text] [doi: 10.1080/10790268.2018.1518124] [Medline: 30207875]
6. Lee H, Lee S, Black I, Salado L, Estrada J, Isla K. Long-Term Impact of Mild Traumatic Brain Injuries on Multiple Functional Outcomes and Epigenetics: A Pilot Study with College Students. Applied Sciences 2020 Jun 16;10(12):4131 [FREE Full text] [doi: 10.3390/app10124131]
7. Dirnberger G, Jahanshahi M. Executive dysfunction in Parkinson's disease: a review. J Neuropsychol 2013 Sep;7(2):193-224. [doi: 10.1111/jnp.12028] [Medline: 24007368]
8. Chastan N, Do MC, Bonneville F, Torny F, Bloch F, Westby GWM, et al. Gait and balance disorders in Parkinson's disease: impaired active braking of the fall of centre of gravity. Mov Disord 2009 Jan 30;24(2):188-195. [doi: 10.1002/mds.22269] [Medline: 18973252]
9. Tomomitsu M, Alonso A, Morimoto E, Bobbio T, Greve J. Static and dynamic postural control in low-vision and normal-vision adults. Clinics 2013 Apr 25;68(4):517-521. [doi: 10.6061/clinics/2013(04)13]
10. Gaerlan, Mary G. The role of visual, vestibular, and somatosensory systems in postural balance (2010). UNLV Theses, Dissertations, Professional Papers, and Capstones 2010:357. [doi: 10.34917/1598677]
11. Ochi F, Abe K, Ishigami S, Otsu K, Tomita H. Trunk motion analysis in walking using gyro sensors. : IEEE; 1997 Oct 30 Presented at: Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 1997; Chicago, IL, USA. [doi: 10.1109/IEMBS.1997.757084]
12. Tsuruoka Y, Shibasaki R. Spectral Analysis in Walking Balance by Elderly Subjects. 2006 Aug 30 Presented at: International Conference of the IEEE Engineering in MedicineBiology Society; 30 Aug 2006; New York p. 5420-5423. [doi: 10.1109/iembs.2006.260446]
13. Tsuruoka M, Tsuruoka Y, Shibasaki R, Yasuoka Y. Spectral Analysis of Walking with Shoes and without Shoes. 2006 Aug 30 Presented at: International Conference of the IEEE Engineering in Medicine and Biology Society; 30 Aug 2006; New York, USA p. 6125-6128. [doi: 10.1109/iembs.2006.4398856]
14. Lee H, Lee S, Salado L, Estrada J, White J, Muthukumar V, et al. Proof-of-Concept Testing of a Real-Time mHealth Measure to Estimate Postural Control During Walking: A Potential Application for Mild Traumatic Brain Injuries. Asian Pac Isl Nurs J 2018;3(4):177-189 [FREE Full text] [doi: 10.31372/20180304.1027] [Medline: 31037266]
15. Kwan RY, Lee D, Lee PH, Tse M, Cheung DS, Thiamwong L, et al. Effects of an mHealth Brisk Walking Intervention on Increasing Physical Activity in Older People With Cognitive Frailty: Pilot Randomized Controlled Trial. JMIR Mhealth Uhealth 2020 Jul 31;8(7):e16596 [FREE Full text] [doi: 10.2196/16596] [Medline: 32735218]
16. Wai AAP, Duc PD, Syin C, Haihong Z. iBEST: Intelligent balance assessmentstability training system using smartphone. 2014 Aug 26 Presented at: 36th Annual International Conference of the IEEE Engineering in MedicineBiology Society; 26 Aug 2014; Chicago, IL p. 3683-3686. [doi: 10.1109/embc.2014.6944422]
17. Rezaie MR, Tahmasebi T, Hassanzadeh A. Can a Smartphone be used for Balance Assessment during Walking in Lower Limb Amputees? Journal of Rehabilitation Sciences & Research 2020;7(2):47-53. [doi: 10.30476/jrsr.2020.84956.1073]

18. Abujrida H, Agu E, Pahlavan K. Smartphone-based gait assessment to infer Parkinson's disease severity using crowdsourced data. 2017 Nov 06 Presented at: IEEE Healthcare InnovationsPoint of Care Technologies (HI-POCT); 06 Nov. 2017; Bethesda, MD p. 208-211. [doi: 10.1109/hic.2017.8227621]
19. SDK Tools release notes. URL: https://developer.android.com/studio/releases/sdk-tools [accessed 2019-08-09]
20. Samsung Galaxy S8 specifications. URL: https://www.samsung.com/global/galaxy/galaxy-s8/specs [accessed 2019-08-09]
21. Android API reference. Android Developers. URL: https://developer.android.com/reference [accessed 2019-08-09]
22. Motion sensors. Android Developers. URL: https://developer.android.com/guide/topics/sensors/sensors_motion [accessed 2021-02-19]
23. Sensors Overview. Android Developers. URL: https://developer.android.com/guide/topics/sensors/sensors_overview [accessed 2021-02-19]
24. Androidplot. URL: http://androidplot.com/ [accessed 2020-06-20]
25. SQLite. URL: https://www.sqlite.org/index.html [accessed 2020-09-10]
26. Cappa P, Jackson JL, Patanè F. Moment Measurement Accuracy of a Parallel Spherical Robot for Dynamic Posturography. IEEE Trans. Biomed. Eng 2010 May;57(5):1198-1208. [doi: 10.1109/tbme.2009.2037807]
27. van Doornik J, Sinkjaer T. Robotic Platform for Human Gait Analysis. IEEE Trans. Biomed. Eng 2007 Sep;54(9):1696-1702. [doi: 10.1109/tbme.2007.894949]
28. Cappa P, Patanè F, Rossi S, Petrarca M, Castelli E, Berthoz A. Effect of changing visual condition and frequency of horizontal oscillations on postural balance of standing healthy subjects. Gait Posture 2008 Nov;28(4):615-626. [doi: 10.1016/j.gaitpost.2008.04.013] [Medline: 18539460]
29. Jürgens R, Boss T, Becker W. Estimation of self-turning in the dark: comparison between active and passive rotation. Exp Brain Res 1999 Oct;128(4):491-504. [doi: 10.1007/s002210050872] [Medline: 10541743]
30. Jürgens R, Nasios G, Becker W. Vestibular, optokinetic, and cognitive contribution to the guidance of passive self-rotation toward instructed targets. Exp Brain Res 2003 Jul;151(1):90-107 [FREE Full text] [doi: 10.1007/s00221-003-1472-y] [Medline: 12740727]
31. Earhart GM, Stevens ES, Perlmutter JS, Hong M. Perception of active and passive turning in Parkinson disease. Neurorehabil Neural Repair 2007;21(2):116-122. [doi: 10.1177/1545968306290674] [Medline: 17312086]
32. Incel O. Analysis of Movement, Orientation and Rotation-Based Sensing for Phone Placement Recognition. Sensors (Basel) 2015 Oct 05;15(10):25474-25506 [FREE Full text] [doi: 10.3390/s151025474] [Medline: 26445046]
33. Sabatini A. Estimating three-dimensional orientation of human body parts by inertial/magnetic sensing. Sensors (Basel) 2011;11(2):1489-1525 [FREE Full text] [doi: 10.3390/s110201489] [Medline: 22319365]
34. Palais B, Palais R, Rodi S. A Disorienting Look at Euler's Theorem on the Axis of a Rotation. Am Math Mon 2009 Dec 01;116(10):892-909. [doi: 10.4169/000298909x477014]
35. Davenport P. Rotations about nonorthogonal axes. AIAA Journal 1973 Jun;11(6):853-857 [FREE Full text] [doi: 10.2514/3.6842]
36. Chumanov ES, Wall-Scheffler C, Heiderscheit BC. Gender differences in walking and running on level and inclined surfaces. Clin Biomech (Bristol, Avon) 2008 Dec;23(10):1260-1268. [doi: 10.1016/j.clinbiomech.2008.07.011] [Medline: 18774631]

Abbreviations

mHealth: mobile health

Edited by L Buis; submitted 08.09.20; peer-reviewed by J Hwang, K Lim; comments to author 06.10.20; revised version received 03.11.20; accepted 05.01.21; published 05.03.21.

Please cite as:
Lee S, Walker RM, Kim Y, Lee H
Measurement of Human Walking Movements by Using a Mobile Health App: Motion Sensor Data Analysis
JMIR Mhealth Uhealth 2021;9(3):e24194
URL: https://mhealth.jmir.org/2021/3/e24194
doi:10.2196/24194
PMID:33666557

Original Paper

An mHealth App (Speech Banana) for Auditory Training: App Design and Development Study

J Tilak Ratnanather[1], DPhil; Rohit Bhattacharya[1*], PhD; Margo B Heston[1*], BS; Joanne Song[1*], BS; Lindsey R Fernandez[1*], PhD; Hong Seo Lim[1*], MS; Seung-Wook Lee[1*], BS; Edric Tam[1*], MS; Sungho Yoo[1*]; Seung-Ho Bae[1*]; Inez Lam[1*], BS; Hyoung Won Jeon[2], MD; Son A Chang[3], PhD; Ja-Won Koo[2], MD, PhD

[1]Center for Imaging Science and Institute for Computational Medicine, Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, United States

[2]Department of Otorhinolaryngology, Seoul National University College of Medicine, Seoul National University Bundang Hospital, Seongnam-si, Republic of Korea

[3]Soree Ear Clinic Rehabilitation Center, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Ja-Won Koo, MD, PhD
Department of Otorhinolaryngology
Seoul National University College of Medicine
Seoul National University Bundang Hospital
82 Gumi-ro 173
Bundang-gu
Seongnam-si, 13620
Republic of Korea
Phone: 82 31 787 7402
Email: speechbanana@jhu.edu

Abstract

Background: With the growing adult population using electronic hearing devices such as cochlear implants or hearing aids, there is an increasing worldwide need for auditory training (AT) to promote optimal device use. However, financial resources and scheduling conflicts make clinical AT infeasible.

Objective: To address this gap between need and accessibility, we primarily aimed to develop a mobile health (mHealth) app called Speech Banana for AT. The app would be substantially more affordable and portable than clinical AT; would deliver a validated training model that is reflective of modern techniques; and would track users' progress in speech comprehension, providing greater continuity between periodic in-person visits. To improve international availability, our secondary aim was to implement the English language training model into Korean as a proof of concept for worldwide usability.

Methods: A problem- and objective-centered Design Science Research Methodology approach was adopted to develop the Speech Banana app. A review of previous literature and computer-based learning programs outlined current AT gaps, whereas interviews with speech pathologists and users clarified the features that were addressed in the app. Past and present users were invited to evaluate the app via community forums and the System Usability Scale.

Results: Speech Banana has been implemented in English and Korean languages for iPad and web use. The app comprises 38 lessons, which include analytic exercises pairing visual and auditory stimuli, and synthetic quizzes presenting auditory stimuli only. During quizzes, users type the sentence heard, and the app provides visual feedback on performance. Users may select a male or female speaker and the volume of background noise, allowing for training with a range of frequencies and signal-to-noise ratios. There were more than 3200 downloads of the English iPad app and almost 100 downloads of the Korean app; more than 100 users registered for the web apps. The English app received a System Usability Scale rating of "good" from 6 users, and the Korean app received a rating of "OK" from 16 users.

Conclusions: Speech Banana offers AT accessibility with a validated curriculum, allowing users to develop speech comprehension skills with the aid of a mobile device. This mHealth app holds potential as a supplement to clinical AT, particularly in this era of global telemedicine.

(JMIR Mhealth Uhealth 2021;9(3):e20890) doi:10.2196/20890

KEYWORDS

speech therapy; mobile phone; computers, handheld; cochlear implants; hearing aids

Introduction

Background

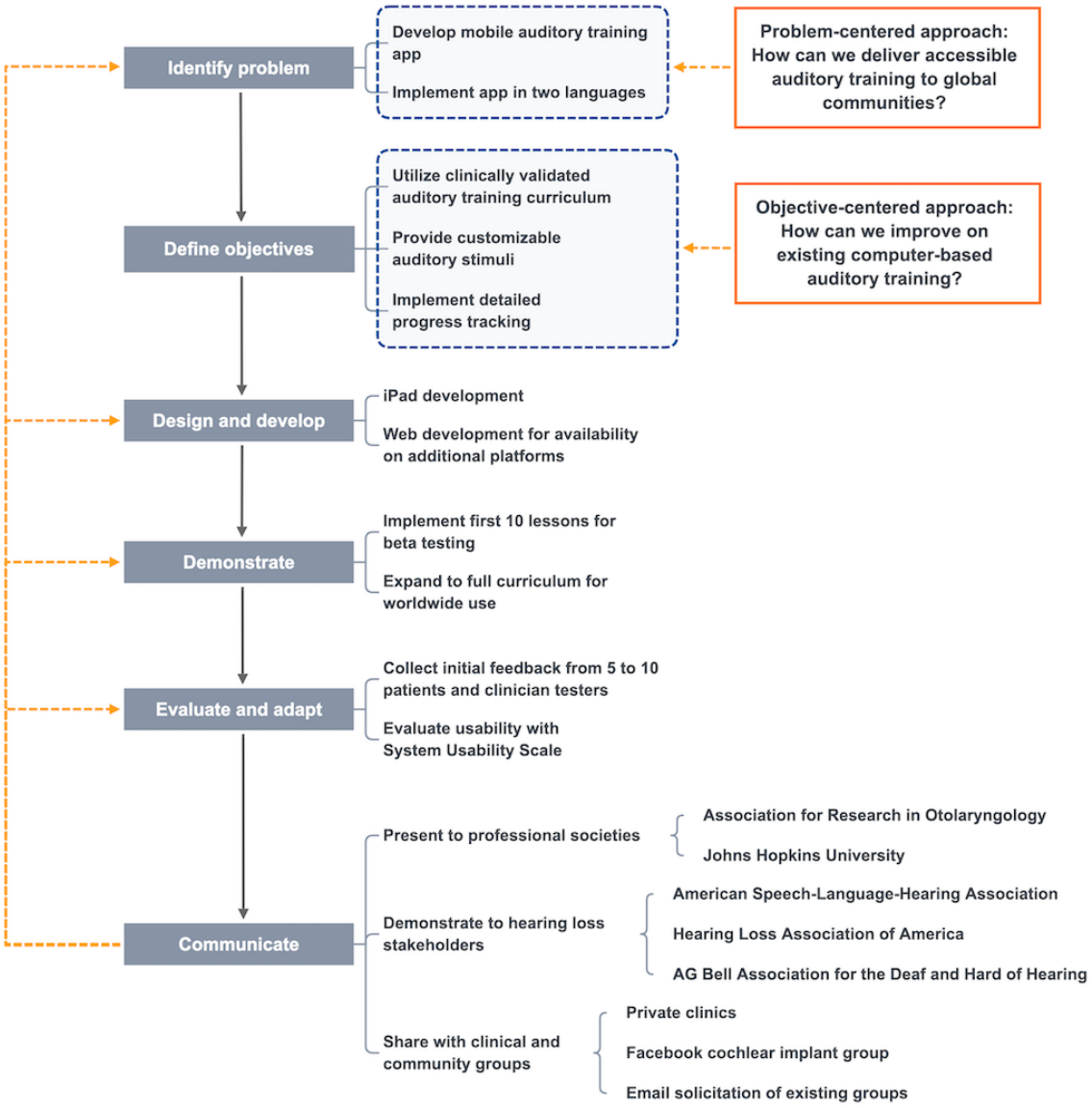
Although auditory prostheses such as hearing aids and cochlear implants have become widely available, several challenges impede adult users from attaining effective hearing rehabilitation. These challenges stem from speech comprehension difficulties, leading to inconsistent device use [1]. Evidence suggests that clinical auditory training (AT) augmented with software programs may improve speech comprehension; however, AT remains globally inaccessible because of financial and clinical time constraints [2-8]. For instance, in the United States, public insurance programs such as Medicare and Medicaid provide insufficient funding to cover the cost of AT [9]. In Korea, children up to 19 years of age may receive AT, but adults are not covered by national health insurance [10]. In the United Kingdom, a lack of AT coverage can leave adult cochlear implant recipients without access to regular sessions outside of annual or semiannual mapping appointments [11,12]. Although clinics attempt to absorb costs through fundraising, sessions may still be prohibitively expensive, leading adults to forgo these services [9]. Even if a patient can afford AT, weekly sessions may require adults to take time away from work, thereby reducing productivity and further exacerbating financial stress [13]. As a result, adults worldwide use their devices without the guidance of a clinician. This motivates the need for accessible AT and for multilingual availability.

One way to address these concerns is to digitize AT and automate the recorded progress. Several groups have developed desktop AT apps, many of which contain proprietary software with management control by clinics [14-16]. All apps use a combination of synthetic (bottom-up) and analytical (top-down) approaches with varying complexity levels; however, they focus primarily on words and phrases, which cultivate skills relevant to the app but not necessarily to daily life. Although desktop-based AT apps are in use, fewer apps are built for mobile devices that are ubiquitous worldwide, such as phones and tablets [17-20]. One free desktop app called i-AngelSound has been redesigned for iPad (Apple Inc) but is not available on other platforms [21,22]. Additional tablet-based apps have emerged for rehabilitating aphasia patients [23,24] and audiological management [25], which indicates a clinical interest in patient-centered mobile solutions. To date, however, there remains a need for an AT app that is appropriate for more advanced adult learning capabilities, is intuitive to use, and is available across mobile platforms.

Objectives

To address this need, Speech Banana was developed as a free mobile health (mHealth) app to provide clinically relevant AT on tablet and web platforms. The name alludes to the shape that speech sound frequencies form when visualized on an audiogram. The Design Science Research Methodology (DSRM) [26] was used to produce the Speech Banana app, guiding development with two problem-based and objective-based processes used as research entry points. The problem-based aim was to provide greater worldwide accessibility to AT, and the objective-based aim was to emulate and improve upon clinical techniques (Figure 1). The first aim was addressed by developing the app on two mobile platforms, which provided portability and flexibility, and by implementing the program in English and Korean languages, which demonstrated a successful expansion to international users. The second aim was addressed by using a clinically validated AT program and by implementing features that speech pathologists are often unable to automate during in-person sessions, such as progress tracking and customizing auditory stimuli. These two processes yielded a mental model of the app wherein users could enhance speech comprehension in their own language and at their own pace. This encourages user engagement through feedback from the app and allows clinicians to deliver targeted in-person AT based on users' at-home progress. The rest of this paper describes the methods for designing, producing, and testing the Speech Banana app; outlines the results of the creation and evaluation processes; and discusses the app's potential to justify insurance funding for AT and for increased spending toward research on computer-based treatment and telemedicine.

Figure 1. Design Science Research Methodology process model for the Speech Banana app.



Methods

Figure 1 summarizes the six components of DSRM and the two nominal processes used as research entry points informing the design.

Training Model

The training model for Speech Banana aimed to (1) implement a validated AT curriculum, (2) provide a training experience similar to in-person clinic visits, (3) expose users to challenging stimuli that would enhance the generalized application of listening skills learned in the app, and (4) demonstrate that the same training model could be implemented across language systems. These objectives were systematically fulfilled through a review of the existing AT literature and clinical interviews.

To ensure a reliable curriculum, training structures common to in-person AT were identified. AT strategies typically implement a combination of drill exercises that focus on analytic skills and a testing regimen focusing on synthetic skills. Analytic drills include repetition of short words or syllables to enhance sound awareness and discrimination skills, whereas synthetic training involves sentences to cultivate sound identification and comprehension through contextual cues [27,28]. Discussions with clinicians at Johns Hopkins University and Woosong University confirmed that the book Auditory Training for the Deaf [29] closely follows this model, making it an ideal source for the content of Speech Banana. It also isolates narrow frequency windows within each lesson, challenging learners to distinguish between similar phonemes. Although the book was published shortly after World War II, its focus on conversational speech skills remains central to clinical AT today [30]. In addition, the book is available in the public domain via the HathiTrust Digital Library, enabling free distribution of Speech Banana without licensing costs. As an aside, the training plan in Auditory Training for the Deaf was modeled on an earlier book, aptly titled Train Your Hearing [31]. With the book's curriculum validated, Auditory Training for the Deaf was adopted as the lesson plan for Speech Banana.

To provide an experience similar to in-person training, auditory and visual cues drawn from clinical references were adapted for an app interface. AT clinicians were observed and interviewed at the Listening Center at Johns Hopkins University,

to identify methods of implementing the analytic and synthetic regimens. Notable training features included allowing lipreading during exercise drills and hiding the clinician's mouth during sentence-based tests. Vocal recording and app structure were accordingly designed to reflect these observations, including visual indicators of analytic material (written words) and providing only auditory cues for synthetic material.

To expose users to challenging stimuli that would enhance the generalized application of listening skills, optional background noise was included in the app. Ambient noise was recorded in a café to provide a multitalker babble that competes with the word- and sentence-based stimuli. This reduces the signal-to-noise ratio per user specifications, creating a progressively more difficult listening experience as the background noise level increases.

To show the training model's applicability across different language systems, a parallel app in Korean was developed. Korean was selected as the second language for Speech Banana because of South Korea's strong national health care system [32] and widespread adoption of smartphones among aging adults [33]. Lesson material was developed at the Department of Speech Language Therapy and Aural Rehabilitation at Woosong University (Daejeon, South Korea) [34].

Auditory Stimuli Recording

According to clinical recommendations, spoken auditory stimuli were recorded using clearly enunciated speech to mimic in-person AT and enable easier phoneme discrimination with hearing devices [35-37]. The recording procedures for English and Korean apps are described below.

For the English version, a Zoom H1 Handy Recorder microphone was used with a pop filter to record the words and sentences in Auditory Training for the Deaf [29]. One male and one female speaker with native English fluency completed the recordings in a soundproofed studio at Johns Hopkins University. Speakers maintained a diction pace slightly slower than standard conversational speech to ensure maximum comprehension. Words were read one at a time in a standard American accent, pausing between each word. Sentences were read in a similar fashion, preserving natural intonation, elisions, and stops and enunciating vowels distinctly. Words and sentences were saved as separate files in waveform audio file format, a lossless audio format. Common words used throughout multiple lessons were recorded and saved once and then copied in lessons as required. Audacity audio editing software was used to splice and normalize tracks, ensuring continuity between recording sessions and eliminating cases of clipping or popping.

For the Korean version, a Neumann TLM-170R microphone (Georg Neumann gmBH) and Adobe Premiere Pro CC were used throughout the recording in a studio at Munhwa Broadcasting Corporation (MBC), a Korean television and radio network station. Two male and two female professional broadcasters from MBC recorded sentences for the Korean app. Each pair of male and female speakers performed half of the lessons. After recording, the sound files were transferred to Johns Hopkins University and subsequently normalized using Audacity.

User Interface

The iPad app was programmed using Xcode, the native language used on the iOS platform. The interface incorporated flat design principles, with few icons, buttons, and navigation links between pages. Contrasting colors and sans serif fonts were used to ensure readability among aging adult users. At the bottom of each page, a Settings button was placed to ensure that features such as speaker sex and background noise level remained accessible throughout the training session. User scores (ie, percentage of correct words and sentences and number of sentences and repetitions) are shown visually as numeric representations. Scores are categorized by lesson, allowing users to compare performance between lessons. To access the app, users must download the software from the Apple App Store onto an iPad.

The web app was programmed in HTML5, JavaScript, and CSS to implement the database and visual elements of the app. The web app was developed using a virtual machine to leverage cloud computing capabilities and allow for app testing on several different operating systems. The web version retained the same color scheme and design principles as the iPad app. User scores (ie, the number of words and sentences correct and number of repetitions) were visualized slightly differently as a bar graph and numeric representation. Scores were categorized by lesson and date of app use, allowing users to compare performance between lessons and over time. To access the web app, users must register with an email address. Upon registration, an identification number is created, and a new user object is created in a MongoDB database to record the quiz performance. Email addresses and passwords were hashed to maintain user security, and no personal health information was stored.

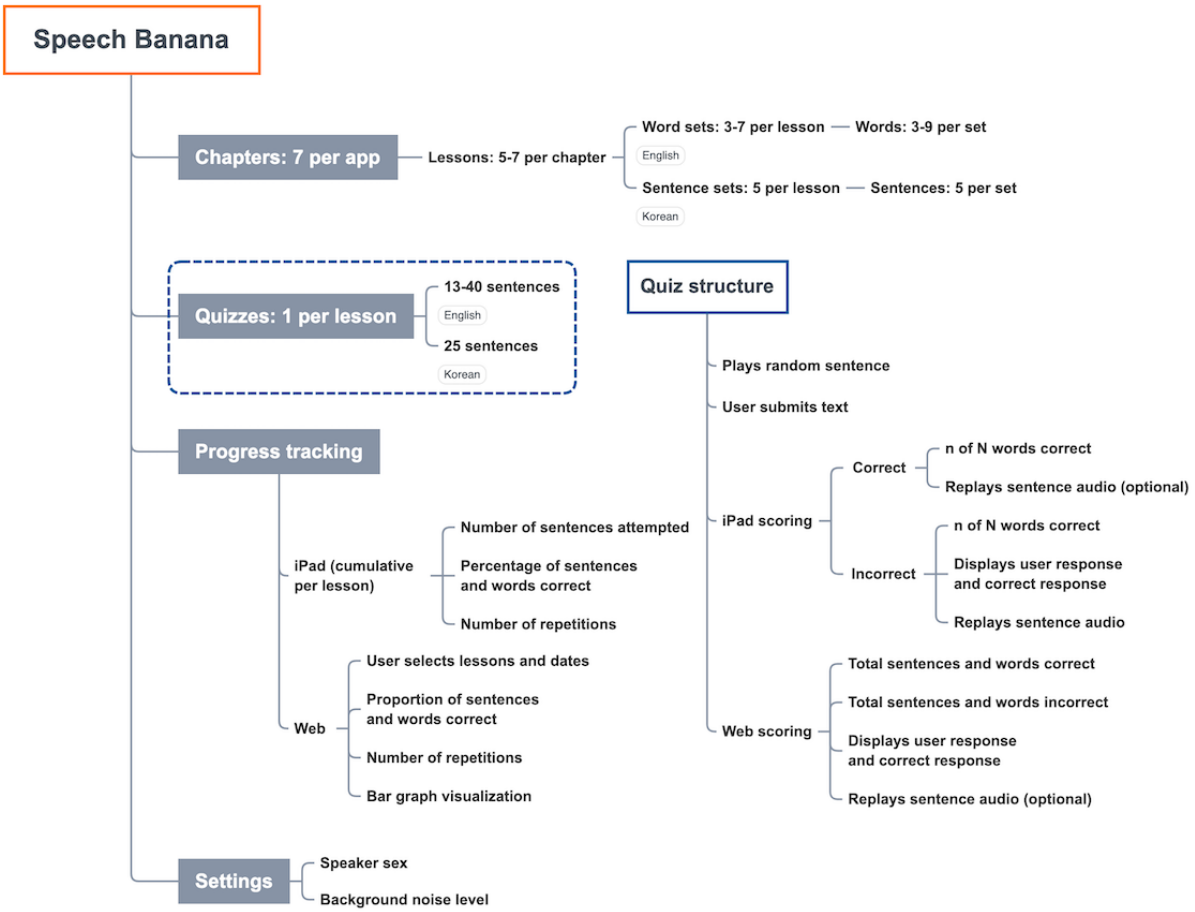
Communication and Evaluation

To evaluate the performance of the iPad and web apps, 2 mechanisms were employed to spread awareness of Speech Banana and gather user feedback: (1) stakeholder feedback via online forums, academic meetings, and community events and (2) a survey based on the System Usability Scale (SUS), a widely used 10-item assessment tool [38]. Using the first evaluation method, updates to the apps were implemented as user comments were logged. For the second evaluation method, past and present users were invited to complete an SUS survey. Scores were averaged, and performance letter grades were generated through comparison against a normalized scale supported by over 500 studies that used the SUS [39].

Results

The English and Korean versions of Speech Banana were deployed for the iPad and web platforms. There were more than 3200 downloads of the English iPad app and almost 100 downloads of the Korean app; more than 100 users registered for the web apps. Figure 2 describes the Speech Banana's structure, function, and usability.

Figure 2. Lesson and quiz structure of the Speech Banana app.



Training Model

The English app follows Auditory Training for the Deaf [29], with 38 lessons grouped into 7 chapters that cover sounds from low to high frequency (Figure 2). This increases the level of difficulty for users with hearing loss at higher frequencies, a common pattern of hearing loss among aging adults [40]. The first 5 lessons provide an introduction to everyday sounds, including numbers, colors, and conversational phrases. The next 16 lessons focus on vowels: lessons 6-11 address low-frequency vowels, lessons 12-17 incorporate vowels that contain high-frequency components, and lessons 18-21 add diphthongs. The remaining 17 lessons focus on low-to-high frequency consonants, pairing unvoiced, higher frequency consonants with their voiced, lower frequency counterparts. This familiarizes users with more difficult phonemes by associating them with phonemes that are easier to distinguish. All speech phonemes are covered during the course of the lesson plan.

Within each lesson, exercises emphasize analytic training, whereas quizzes incorporate synthetic training. Exercises include visual and auditory stimuli, displaying written words to accompany the sounds of single words and 2-word phrases. Users may repeat these stimuli as needed to practice discrimination and identification skills. After completing a lesson, the app directs users to an interactive quiz that plays auditory-only, sentence-based stimuli (Figure 2, Quizzes). Users may replay sentences before typing their answers into a text box. Throughout the program, lesson word banks increase in size, challenging the user to distinguish larger sets of information. Similarly, quizzes increase in length, providing more sentences and referring to words of earlier lessons to reinforce the material.

The Korean app differs from the English version to accommodate for phonetic and syntactic differences between the two languages and to retain lessons' progression from low- to high-frequency phonemes [41]. In addition, although the English app contains word and sentence stimuli, the Korean lessons comprise only sentences. This is because existing Korean training and evaluation services use word-based stimuli; however, they lack sentence-based training; therefore, the app was designed to address this training gap. The first 5 lessons comprise brief phrases and sentences, including nouns and verbs, colors, numbers, and common expressions. Following the initial lessons, vowels and consonants are covered in a similar order to the English version. Lessons 6-12 focus on low-frequency vowels, and lessons 13-18 focus on voiced consonants progressing from low to high frequencies. Lessons 19-23 focus on high-frequency vowels, and lessons 24 to 28 focus on unvoiced consonants from low to high frequencies. Lessons 29-33 provide contrasts between vowels and consonants. Lessons 34-35 focus on prosody, such as nursery rhymes, traditional songs, and well-known sports cheers, which are used to encourage engagement with the app. Lessons 36-38 feature advanced auditory discrimination tasks in the form of nonsense sentences. These remove the aid of context clues, helping to

develop discrimination acuity and strengthen auditory working memory.

Auditory Stimuli Recording

Approximately 1170 pairs of words and 1000 pairs of sentences were recorded for the English app. Audio files from lessons 1-10 are freely available to the public on Freesound, an online sound sample repository [42]. These recordings have been used in online games, videos, and educational software. For the Korean app, 950 pairs of sentences were recorded. As the programs include a large number of lossless audio files, the app size is 1.75 GB for the English app and 0.94 GB for the Korean app. This retains high-fidelity playback but occupies considerable storage space on mobile devices.

Both apps showed mild variability in speaker volume or speed. Sporadically, throughout the English language app, the speaker volume varies; meanwhile, in the Korean app, the pair of speakers in the first 19 lessons maintained a slightly quieter tone and faster pace than the second pair. These volume level differences create a more challenging listening experience for new users, and they may be attenuated through additional normalization or rerecording in future app versions.

User Interface

The apps may be accessed from the Speech Banana web portal [43]. Users may proceed from the portal to the Apple App Store to download the iPad app, or may log into the web app, Google authorization or Naver authorization for the Korean app. Upon entering the app, users encounter the main page that contains motivation for Speech Banana, instructions for use, and links to lessons and quizzes. Curriculum pages have three elements: a header to select speaker sex and choose word sets within a lesson, the sidebar menu to navigate between sections of the app, and the main content space to complete lessons and quizzes.

Lessons contain word or sentence stimuli divided into sets that present materials in manageable learning units. Stimuli are displayed visually in large buttons that users may press to play the associated audio file. At any point in a lesson, users may navigate directly to the corresponding quiz using a gray button on the right side of the page. As users progress through the quiz, scores are updated in real time and are viewable beneath the input bar (web app only) or in the progress section of the app.

During quizzes, users press a button at the center of the page to play a sentence; in response, the app randomly generates a number, searches for the appropriately labeled audio, and plays the file. To ensure that all sentences within the sentence bank are exhaustively selected, numbers are drawn without replacement, and the random function is reset after completing the quiz. Users may play the sentence several times before typing an estimate of what they heard. Input is scored based on syntax and spelling, but punctuation is disregarded. If the written response does not match the auditory stimulus, users' input and the correct sentences are displayed alongside users' score, which indicates the number of words typed correctly. The auditory stimulus may then be replayed with the visual stimulus to improve the identification skills.

Background noise and speaker sex may be user-specified throughout the app (Figure 2). When selected, background noise plays continuously at the volume specified by an onscreen slide bar, and the maximum volume is determined by the device's output capacity. During quizzes, users of the iPad app may select the speaker sex, whereas the web app randomizes speakers using a uniformly distributed random variable.

Progress Tracking

Scoring on the iPad app is recorded separately for each lesson and displayed on the progress page in a list format (Figure 2). Web app scores are similarly recorded but are visualized with additional tabulation and a dynamic bar chart. Users can display results within a specified period and may filter any of the metrics by lessons or chapters.

Evaluation and Communication

During development, 2 rounds of feedback were gathered, which helped solidify the user interface and expand to web development. After app publication, communication was primarily conducted among English-speaking audiences and included academic presentations at professional meetings, consultations with clinicians, and demonstrations at community events.

Preliminary Feedback

The English version of the iPad app was first developed as an alpha version, consisting of the first 10 lessons and deployed on TestFlight via Apple Store Connect. A total of 10 testers and clinicians were invited to use the app and provide feedback on app navigation, visual esthetic, and auditory levels. After an initial round of feedback, the remaining 28 lessons were implemented.

User Feedback

Once the full curriculum was published for iPad use on Apple Store, a continuous stream of feedback was received from the Cochlear Implant Experiences Facebook group, a community page that sees traffic from more than 30,000 members at the time of publication. One salient point of feedback was that people enjoyed the large screen format of a tablet; however, prospective users were unable to access the app because they did not own an iPad. A literature review confirmed this sentiment; although the majority of adults in both advanced and emerging economies use the internet on smartphones, they are much less likely to use computers or tablets [44]. To maximize the accessibility of Speech Banana on any smart device, the team moved forward with web development in English and Korean. The web versions were published in December 2019, and account registration on Google and Naver was added in May 2020. In addition to web development, user comments spurred the creation of online instructions and a frequently asked questions section on the Speech Banana web portal [43]. This helped address common questions regarding app navigation, volume control, and hardware compatibility.

SUS Score

Table 1 shows the Speech Banana usability scores from past and present users of the app with letter grades derived from a normalized scale [39].

Table 1. Speech Banana usability scores. Letter grades are derived from a normalized scale.

App version	System Usability Scale score, mean (SD)	Grade
English	79.17 (15.39)	B− (Good)
Korean	65.63 (17.99)	D (OK)

The English app received 6 responses to the SUS survey; 4 respondents used the iPad app, 1 used the web app, and 1 used both apps. The Korean app achieved greater SUS participation with 16 scores; 5 from iPad users and 11 from web app users. On the basis of the normalized SUS grades, the English apps achieved good usability, whereas the Korean version performed adequately. The survey results revealed that an initial but brief training session could be implemented to improve user satisfaction and comfort with the apps.

Communication

Following the public release of the English iPad app, the app was presented at academic meetings, including the Association for Research in Otolaryngology Annual Meeting, which is the premier conference for research in the auditory and vestibular sciences [45,46]. Talks and posters were additionally presented to clinical stakeholders in the hearing loss community, such as the American Speech-Language and Hearing Association [47]. Finally, updates were shared with clinical and community groups via Johns Hopkins University publications, social media, and electronic mailing lists [48].

Discussion

Principal Findings

This study describes the design, development, and assessment of Speech Banana, an mHealth app that responds to a pressing global health need for accessible and effective hearing health care. The app digitizes a validated clinical model that uses word-based exercises to improve users' familiarity with targeted speech sounds and that delivers sentences-based quizzes to help users generalize their training to conversational listening. Additional features such as tracking enable users to visualize progress with speech comprehension. To make the app available to populations worldwide with multiple languages and access via different electronic devices, Speech Banana was developed in two languages and hosted on two digital platforms. Continuous feedback was used to improve the app, and a pilot usability study rated the mHealth app as adequate to good. These results suggest that the app is successful in providing a usable digital service, and it shows promise as an effective AT option that minimizes the need to pay for or attend clinical rehabilitation sessions. Since its initial worldwide release, Speech Banana has undergone scholarly review, highlighting some of its more successful features and identifying others that would improve the use of the app. Some features have been incorporated into software updates, whereas other features will require future work or technological advancements to fully implement.

With the World Health Organization estimating 466 million people with disabling hearing loss, the lack of accessible AT has become a substantial public health need [49]. Many hearing devices now connect to portable sound amplification products, including smartphones, which have achieved 99% global market penetration [50,51]. This presents an unprecedented opportunity to distribute AT to more people with hearing loss than ever before and to enable users to manage their own hearing health care. The Speech Banana team sought to leverage this opportunity, resulting in an app that is grounded in clinically supported training regimens, implemented in a language with substantially different syntax, and supported across mobile platforms [52].

In designing and executing the Speech Banana app, a 2-pronged design science research approach was chosen, to improve the availability and efficacy of AT worldwide. Traditionally, design science research originates from a single point of entry, which informs subsequent objectives and implementation of the project [26]. From the outset, however, issues were identified both with AT availability and with other AT apps' delivery and progress tracking methods. The former issue called for the development of a mobile artifact that may be expanded to multiple languages. After initial development and feedback, it became clear that accessibility would not only require an app developed in multiple languages but that the mobile platform itself would significantly impact users' barriers to entry [44]. With this information, web development was judged to be the most appropriate; therefore, the app could be used on any mobile device with internet access. The latter issue required dedicated attention to the AT curriculum, separate from any technical considerations. Although Auditory Training for the Deaf [29] is out of print, evidence shows that its structure (analytic lessons and synthetic quizzes) is effective for developing speech comprehension skills [27,28]. Furthermore, analytic and synthetic skill building and transition from low- to high-frequency phonemes makes the lesson plan easily adaptable to any language, satisfying both the accessibility and clinical relevance aims. Although the dual-pronged problem statement was nontraditional per established DSRM, it made for a more robust artifact that has strong expansion potential.

Since the release of Speech Banana, additional apps have entered the AT market, and an academic review of these apps has followed [17]. In this review, Olson et al [17] performed a survey of more than 200 mobile apps for AT and deemed 5 iPad apps, including Speech Banana, appropriate for a detailed review. Speech Banana meets the 5 characteristics that were expected of an AT app: (1) it provides feedback through real-time scoring and gives the user an opportunity to repeat the test stimulus, (2) it uses a large training corpus of words and sentences, (3) it trains users on specific phonemes across the speech frequency space, (4) it employs analytic and synthetic processing, and (5) it tracks user performance via a progress report [53]. With these criteria met, Speech Banana is competitive among its peers in the AT app space.

Limitations and Future Work

Although Speech Banana addresses the original aims, it should be noted that there are several limitations that merit addressing in future development of the app, or that may require technological advances. First, using lossless files for auditory stimuli leads to a large total app size, which consumes a substantial portion of the available memory space on current iPads. These files are necessary to ensure clear audio for users learning to hear; however, it may deter some users from downloading the app. As mobile storage space continues to increase, space limitations may become less of a concern. For users who do not own an iPad, use of the web app may be infeasible with current internet connectivity; however, as internet access becomes increasingly ubiquitous, this limitation will diminish, enabling easier distribution of Speech Banana and mobile AT more broadly. Second, although Speech Banana's stimuli are of high quality, the auditory output may be limited by the capabilities of iPad or smartphone speakers. This may be mitigated by using headphones, direct audio input, or Bluetooth connectivity. Third, users mentioned volume inconsistency across the sound files. This may be reduced by reprocessing the raw audio files or through further normalization to account for the recording of batch effects. Fourth, older users found it difficult to type responses in the quiz windows. This can be significantly alleviated by using a stylus pencil that works with the newer iPad models.

Apart from basic functionality limitations, feedback has suggested several feature additions that could improve user experience, including phonemic progress tracking, adaptive quizzes, and additional languages. Phonemic progress tracking would significantly advance the app's integration with periodic in-person AT. Word- and sentence-based achievement used in the current version of the app does not inform users of frequency or timing cues that require additional training. If the app tracked phoneme comprehension, Speech Banana could better aid targeted clinical interventions or guide the calibration of hearing devices. One potential method could employ cloud computing to classify and interpret trends in phoneme comprehension errors. In addition to improving speech comprehension, phonemic tracking could be used to train the vocalization of challenging phonemes, enhancing speech production. Adaptive quizzes can improve user engagement and prevent frustration by adjusting the difficulty based on previous performance. This may elongate the session length, increasing the potential efficacy of the app. Language additions have been requested for several languages, including British English, French, German, Turkish, Arabic, Spanish, Hindi, Tamil, and Sinhalese. These would enhance global accessibility, but developing the app is time consuming, particularly with a small team. Implementation may be accelerated by using publicly available speech corpora such as the British English Speech Corpus, which is frequently used in speech recognition research. To ensure phonemic consistency across languages, speech and language clinicians should be invited to collaborate on curriculum development.

Finally, although Speech Banana uses a curriculum that conforms to clinical AT methods, the app itself has not been validated for efficacy, skill generalization, computer-based AT compliance, and training durability [1,54-56]. To address this, a randomized clinical trial should be conducted to evaluate the use of the app on different mobile devices, comparing efficacy among patients with and without access to supplemental in-person AT [52,57]. As clinical interest in telemedicine expands, this could serve as evidence to justify increased private and public insurance funding toward AT availability and toward the development of AT technologies.

As the global aging population grows, the incidence of hearing loss continues to rise, requiring that options for AT are made accessible to mitigate or forestall cognitive decline [58]. Isolation because of the recent COVID-19 pandemic has demonstrated the necessity of telemedicine advancements and has provided evidence that remote health care can be effective with careful design and implementation. Continued use and evaluation of the Speech Banana app will help us understand the efficacy of computer-based AT and will aid the development of coordinated care that integrates clinical time with home practice. Through the current app and future extensions, Speech Banana may be an effective tool to broaden access to hearing health care worldwide.

Acknowledgments

The authors dedicate this paper to the memory of John Kim Niparko, MD, the latest in a long line of otologist-scientists going back to Viktor Urbantschitsch who advanced the field of auditory rehabilitation for people with hearing loss. He will forever "live in the words heard by his patients."

The authors would like to thank Joshua Langfus for providing the male speaker stimuli in the English version. The authors would also like to thank Kim Bum Do, Hur Il Hoo, Lee Jaeeun, and Cha Yerin for lending their voices; Yun Hong Il for recording direction; and MBC for providing time and facilities toward the Korean version. In addition, the authors would like to thank Dr Howard Francis and Ms Kristin Ceh of the Johns Hopkins Listening Center (now Johns Hopkins Cochlear Implant Center) for their encouragement. The authors would like to thank Dr Youngser Park and Anthony Kolasny for their assistance with the Korean app and web app, respectively. Finally, the authors would like to thank the people with hearing loss worldwide who provided feedback on early versions of the app, as their suggestions were crucial to ensure user satisfaction and engagement with computer-based AT.

Authors' Contributions

Concept was developed by JR. Design was done by RB, MH, JS, HL, SL, and SC. Implementation was carried out by RB, LF, HSL, and SWL. Deployment was done by RB, ET, SY, and SB. Testing and dissemination were carried out by JR, MH, JS, LF, HL, and IL. Audio was developed by MH, HL, and JK. Data analysis was carried out by SB, MH, and HJ. The manuscript was written by JR, MH, JS, SL, SB, SC, and JK.

Conflicts of Interest

None declared.

References

1. Henshaw H, McCormack A, Ferguson MA. Intrinsic and extrinsic motivation is associated with computer-based auditory training uptake, engagement, and adherence for people with hearing loss. Front Psychol 2015;6:1067 [FREE Full text] [doi: 10.3389/fpsyg.2015.01067] [Medline: 26300796]
2. Baungaard LH, Sandvej MG, Krøijer JS, Hestbæk MK, Samar CF, Percy-Smith L, et al. Auditory verbal skills training is a new approach in adult cochlear implant rehabilitation. Dan Med J 2019;66(3) [FREE Full text] [Medline: 30864546]
3. Magro I, Emmett SD, Saunders J. Cost-effectiveness of CI in developing countries. Curr Opin Otolaryngol Head Neck Surg 2018;26(3):190-195. [doi: 10.1097/MOO.0000000000000451] [Medline: 29528875]
4. Olson AD. Options for auditory training for adults with hearing loss. Semin Hear 2015;36(4):284-295 [FREE Full text] [doi: 10.1055/s-0035-1564461] [Medline: 27587915]
5. Plant G, Bernstein CM, Levitt H. Optimizing performance in adult cochlear implant users through clinician directed auditory training. Semin Hear 2015;36(4):296-310 [FREE Full text] [doi: 10.1055/s-0035-1564460] [Medline: 27587916]
6. Stropahl M, Besser J, Launer S. Auditory training supports auditory rehabilitation: a state-of-the-art review. Ear Hear 2020;41(4):697-704. [doi: 10.1097/AUD.0000000000000806] [Medline: 31613823]
7. Tao KFM, Brennan-Jones CG, Capobianco-Fava DM, Jayakody DMP, Friedland PL, Swanepoel DW, et al. Teleaudiology services for rehabilitation with hearing aids in adults: a systematic review. J Speech Lang Hear Res ;61(7):1831-1849. [doi: 10.1044/2018_JSLHR-H-16-0397] [Medline: 29946688]
8. Bierbaum M, McMahon CM, Hughes S, Boisvert I, Lau AYS, Braithwaite J, et al. Barriers and facilitators to cochlear implant uptake in Australia and the United Kingdom. Ear Hear 2020;41(2):374-385. [doi: 10.1097/AUD.0000000000000762] [Medline: 31356385]
9. Sorkin DL. Cochlear implantation in the world's largest medical device market: utilization and awareness of cochlear implants in the United States. Cochlear Implants Int 2013;14 Suppl 1:S4-12 [FREE Full text] [doi: 10.1179/1467010013z.00000000076] [Medline: 23453146]
10. Byeon H. Development trends of online-based aural rehabilitation programs for children with cochlear implant coping with the fourth industrial revolution and implication in speech-language pathology. Int J Adv Comp Sci App 2019;10(10). [doi: 10.14569/ijacsa.2019.0101005]
11. Raine C. Cochlear implants in the United Kingdom: awareness and utilization. Cochlear Implants Int ;14 Suppl 1:S32-S37 [FREE Full text] [doi: 10.1179/1467010013Z.00000000077] [Medline: 23453150]
12. Raine C, Atkinson H, Strachan DR, Martin JM. Access to cochlear implants: time to reflect. Cochlear Implants Int 2016;17 Suppl 1:42-46. [doi: 10.1080/14670100.2016.1155808] [Medline: 27099110]
13. Reis M, Boisvert I, Beedell E, Mumford V. Auditory training for adult cochlear implant users: a survey and cost analysis study. Ear Hear 2019;40(6):1445-1456. [doi: 10.1097/AUD.0000000000000724] [Medline: 30998544]
14. Boothroyd A. Adapting to changed hearing: the potential role of formal training. J Am Acad Audiol 2010;21(9):601-611. [doi: 10.3766/jaaa.21.9.6] [Medline: 21241648]
15. Levitt H, Oden C, Simon H, Noack C, Lotze A. Entertainment overcomes barriers of auditory training. Hearing J 2011;64(8):40. [doi: 10.1097/01.hj.0000403510.80465.7b]
16. Sweetow RW, Sabes JH. Auditory training and challenges associated with participation and compliance. J Am Acad Audiol 2010;21(9):586-593. [doi: 10.3766/jaaa.21.9.4] [Medline: 21241646]
17. Olson A, Williams R, Livingston E, Futscher C. Review of auditory training mobile apps for adults with hearing loss. Perspect ASHA SIGs 2018;3(7):12-23. [doi: 10.1044/persp3.sig7.12]
18. Paglialonga A, Nielsen AC, Ingo E, Barr C, Lévesque AL. eHealth and the hearing aid adult patient journey: a state-of-the-art review. Biomed Eng Online ;17(1):101 [FREE Full text] [doi: 10.1186/s12938-018-0531-3] [Medline: 30064497]
19. Paglialonga A, Tognola G, Pinciroli F. Apps for hearing science and care. Am J Audiol 2015;24(3):293-298. [doi: 10.1044/2015_AJA-14-0093] [Medline: 26649533]
20. Paglialonga A, Tognola G, Pinciroli F. Apps for hearing healthcare. Stud Health Technol Inform 2015;210:666-668. [Medline: 25991233]
21. Fu QJ, Galvin JJ. Perceptual learning and auditory training in cochlear implant recipients. Trends Amplif 2007;11(3):193-205 [FREE Full text] [doi: 10.1177/1084713807301379] [Medline: 17709574]
22. Fu QJ, Galvin III JJ. Auditory training for cochlear implant patients. Switzerland: Springer; 2012:257-278.

23. Roches CAD, Balachandran I, Ascenso EM, Tripodis Y, Kiran S. Effectiveness of an impairment-based individualized rehabilitation program using an iPad-based software platform. Front Hum Neurosci 2014;8:1015 [FREE Full text] [doi: 10.3389/fnhum.2014.01015] [Medline: 25601831]
24. Gerber SM, Schütz N, Uslu AS, Schmidt N, Röthlisberger C, Wyss P, et al. Therapist-guided tablet-based telerehabilitation for patients with aphasia: proof-of-concept and usability study. JMIR Rehabil Assist Technol ;6(1):- [FREE Full text] [doi: 10.2196/13163] [Medline: 31025946]
25. Larrosa F, Rama-Lopez J, Benitez J, Morales JM, Martinez A, Alañon MA, et al. Development and evaluation of an audiology app for iPhone/iPad mobile devices. Acta Otolaryngol 2015;135(11):1119-1127. [doi: 10.3109/00016489.2015.1063786] [Medline: 26144548]
26. Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S. A design science research methodology for information systems research. J Manag Inf Syst ;24(3):45-77. [doi: 10.2753/MIS0742-1222240302]
27. Erber NP. Auditory training. Washington DC: Alexander Graham Bell Association for the Deaf; 1982.
28. Erber NP. Communication therapy for hearing-impaired adults. Australia: Clavis Publishing; 1988.
29. Whitehurst MW, Monsees EK. Auditory training for the deaf. Washington DC: The Volta Bureau; 1952.
30. Bergman M. On the origins of audiology: American wartime military audiology (Audiology Today Monograph No. 1). McLean, VA: American Academy of Audiology; 2002. URL: https://www.audiology.org/sites/default/files/audiologytoday/2002ATJanSpecial.pdf [accessed 2021-02-24]
31. Whitehurst MW. Train your hearing. Washington DC: The Volta Bureau; 1947.
32. World Health Organization. Republic of Korea health system review. Manila: WHO Regional Office for the Western Pacific; 2015.
33. Silver L, Smith A, Johnson C, Taylor K, Jiang J, Monica A, et al. Use of smartphones and social media is common across most emerging economies. Pew Research Center 2019 [FREE Full text]
34. Chang S, Jeon J, Koo J, Oh S, Ratnanather T. Development of listening evaluation and practice tool in a format of mobile application. In: Proceedings of the Seoul International Conference on Speech Sciences. 2017 Presented at: Seoul International Conference on Speech Sciences; 2017; Seoul.
35. Picheny MA, Durlach NI, Braida LD. Speaking clearly for the hard of hearing I: intelligibility differences between clear and conversational speech. J Speech Hear Res 1985;28(1):96-103. [doi: 10.1044/jshr.2801.96] [Medline: 3982003]
36. Picheny MA, Durlach NI, Braida LD. Speaking clearly for the hard of hearing. II: Acoustic characteristics of clear and conversational speech. J Speech Hear Res 1986;29(4):434-446. [doi: 10.1044/jshr.2904.434] [Medline: 3795886]
37. Picheny MA, Durlach NI, Braida LD. Speaking clearly for the hard of hearing. III: an attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech. J Speech Hear Res 1989;32(3):600-603. [Medline: 2779204]
38. Friesen EL. Measuring AT usability with the modified System Usability Scale (SUS). Stud Health Technol Inform 2017;242:137-143. [Medline: 28873790]
39. Brooke J. SUS: a retrospective. J usability stud. 2013. URL: https://uxpajournal.org/sus-a-retrospective/ [accessed 2021-02-24]
40. Lin FR. Hearing loss and cognition among older adults in the United States. J Gerontol A Biol Sci Med Sci 2011;66(10):1131-1136 [FREE Full text] [doi: 10.1093/gerona/glr115] [Medline: 21768501]
41. Kim BJ, Chang S, Yang J, Oh S, Xu L. Relative contributions of spectral and temporal cues to Korean phoneme recognition. PLoS One 2015;10(7):- [FREE Full text] [doi: 10.1371/journal.pone.0131807] [Medline: 26162017]
42. Heston M. 2013. URL: http://freesound.org/people/margo_heston [accessed 2021-02-24]
43. Speech Banana Team. 2015. URL: http://speechbanana.jhu.edu [accessed 2020-11-01]
44. Taylor K, Silver L. Smartphone ownership is growing rapidly around the world, but not always equally. Pew Research Center 2019;5. [doi: 10.4135/9781412963947.n376]
45. Bhattacharya R, Heston M, Song J, Fernandez L, Ratnanather T. Challenges in designing and developing an auditory training iPad app for adults with hearing loss. In: Proceedings of the 38th Annual MidWinter Meeting of Association for Research in Otolaryngology. 2015 Presented at: 38th Annual MidWinter Meeting of Association for Research in Otolaryngology; 2015; Baltimore, MD p. 80.
46. Ratnanather T. Speech Banana - a modular iPad app for auditory training for adults with hearing loss. In: Proceedings of the 38th Annual Midwinter Meeting of the Association for Research in Otolaryngology. 2015 Presented at: 38th Annual Midwinter Meeting of the Association for Research in Otolaryngology; 2015; Baltimore, MD p. 397.
47. Ratnanather JT. Speech Banana: a mobile platform for auditory training for adults with hearing loss. ASHA KIRAN. 2015. p. 11-14 URL: https://www.sites.google.com/site/asianindiancaucusasha1/newsletter [accessed 2021-03-10]
48. Regan MB. Hear, hear! An app for cochlear implant users. JHU Engineering. 2014. URL: https://engineering.jhu.edu/magazine/2014/05/hear-hear-app-cochlear-implant-users/ [accessed 2021-02-24]
49. Deafness and hearing loss. World Health Organization. 2018. URL: https://wwwwhoint/news-room/fact-sheets/detail/deafness-and-hearing-loss [accessed 2020-11-01]
50. Swanepoel DW. eHealth technologies enable more accessible hearing care. Semin Hear 2020;41(2):133-140. [doi: 10.1055/s-0040-1708510] [Medline: 32269417]

51. Reed N, Korczak P, Polyak N, Lin F. Comparison of PSAPs and hearing aids: the role of affordable and accessible devices in hearing healthcare. Innovation in Aging. (suppl_1) 2017;1:1386. [doi: 10.1093/geroni/igx004.5104] [Medline: PMC6183496]
52. Ferguson M, Maidment D, Henshaw H, Heffernan E. Evidence-based interventions for adult aural rehabilitation: that was then, this is now. Semin Hear 2019;40(1):68-84 [FREE Full text] [doi: 10.1055/s-0038-1676784] [Medline: 30728650]
53. Watson CS, Miller JD, Kewley-Port D, Humes LE, Wightman FL. Training listeners to identify the sounds of speech: I. A review of past studies. Hear J ;61(9):26 [FREE Full text] [doi: 10.1097/01.hj.0000339502.52055.d8] [Medline: 20098630]
54. Bronus K, El Refaie A, Pryce H. Auditory training and adult rehabilitation: a critical review of the evidence. Glob J Health Sci ;3(1). [doi: 10.5539/gjhs.v3n1p49]
55. Henshaw H, Ferguson MA. Efficacy of individual computer-based auditory training for people with hearing loss: a systematic review of the evidence. PLoS One 2013;8(5):- [FREE Full text] [doi: 10.1371/journal.pone.0062836] [Medline: 23675431]
56. Henshaw H, Ferguson M. Assessing the benefits of auditory training to real-world listening: identifying appropriate and sensitive outcomes. In: Proceedings of the International Symposium on Auditory and Audiological Research. 2013 Presented at: International Symposium on Auditory and Audiological Research; 2013; Nyborg, Denmark p. 45-52.
57. Hixon B, Chan S, Adkins M, Shinn JB, Bush ML. Timing and impact of hearing healthcare in adult cochlear implant recipients: a rural-urban comparison. Otol Neurotol 2016;37(9):1320-1324 [FREE Full text] [doi: 10.1097/MAO.0000000000001197] [Medline: 27636389]
58. Nkyekyer J, Meyer D, Pipingas A, Reed NS. The cognitive and psychosocial effects of auditory training and hearing aids in adults with hearing loss. Clin Interv Aging 2019;14:123-135 [FREE Full text] [doi: 10.2147/CIA.S183905] [Medline: 30666098]

Abbreviations

AT: auditory training
DSRM: Design Science Research Methodology
MBC: Munhwa Broadcasting Corporation
mHealth: mobile health
SUS: System Usability Scale

Edited by L Buis; submitted 21.06.20; peer-reviewed by M Prietula, O El-Gayar; comments to author 31.07.20; revised version received 31.10.20; accepted 10.02.21; published 15.03.21.

Please cite as:
Ratnanather JT, Bhattacharya R, Heston MB, Song J, Fernandez LR, Lim HS, Lee SW, Tam E, Yoo S, Bae SH, Lam I, Jeon HW, Chang SA, Koo JW
An mHealth App (Speech Banana) for Auditory Training: App Design and Development Study
JMIR Mhealth Uhealth 2021;9(3):e20890
URL: https://mhealth.jmir.org/2021/3/e20890
doi:10.2196/20890
PMID:33720025

Original Paper

Smartphone Medical App Use and Associated Factors Among Physicians at Referral Hospitals in Amhara Region, North Ethiopia, in 2019: Cross-sectional Study

Gizaw Hailiye Teferi[1], BSc, MPH; Binyam Cheklu Tilahun[2], BSc, MPH, MSc, PhD; Habtamu Alganeh Guadie[3], BSc, MPH; Ashenafi Tazebew Amare[4], MD

[1]Department of Health Informatics, College of Medicine and Health Science, Debre-Markos University, Debre Markos, Ethiopia

[2]Department of Health Informatics, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia

[3]School of Public Health, College of Medicine and Health Science, Bahir Dar University, Bahir Dar, Ethiopia

[4]Department of Pediatrics, College of Medicine and Health Science, University of Gondar, Gondar, Ethiopia

Corresponding Author:
Gizaw Hailiye Teferi, BSc, MPH
Department of Health Informatics
College of Medicine and Health Science
Debre-Markos University
PO Box: 269
Debre Markos
Ethiopia
Phone: 251 921493560
Email: ghailiye463@gmail.com

Abstract

Background: Information in health care is rapidly expanding and is updated very regularly, especially with the increasing use of technology in the sector. Due to this, health care providers require timely access to the latest scientific evidence anywhere. Smartphone medical apps are tools to access the latest reputable scientific evidence in the discipline. In addition, smartphone medical apps could lead to improved decision making, reduced numbers of medical errors, and improved communication between hospital medical staff.

Objective: The aim of this study was to assess smartphone medical app use and associated factors among physicians working at referral hospitals of the Amhara region, Ethiopia.

Methods: An institution-based cross-sectional study design was conducted among physicians working at 5 referral hospitals in the Amhara region, Ethiopia, from February 5 to May 27, 2019. A simple random sampling method was used to select 423 physicians. A self-administered questionnaire was used to collect the data and analyzed using SPSS, version 21 (IBM Corp). Binary and multivariable logistic regression analysis was performed to assess factors associated with smartphone medical app use among physicians. A value of $P<.05$, corresponding to a 95% CI, was considered statistically significant. The validity of the questionnaire was determined based on the view of experts and the reliability of it obtained by calculating the value of Cronbach alpha ($\alpha=.78$)

Results: In this study, most of the 417 respondents (375, 89.9%) had medical apps installed on their smartphones. Of those 375 respondents, 264 (70.4%) had used medical apps during clinical practice. The medical apps most commonly used by the respondents were UpToDate, Medscape, MedCalc, and Doximity. According to multivariable logistic regression analysis, attitude (adjusted odds ratio [AOR] 1.64, 95% CI 1.05-2.55), internet access (AOR 2.82, 95% CI 1.75-4.54), computer training (AOR 1.71, 95% CI 1.09-2.67), perceived usefulness of the app (AOR 1.64, 95% CI 1.05-2.54), information technology support staff (AOR 2.363, 95% CI 1.5-3.08), and technical skill (AOR 2.52, 95% CI 1.50-4.25) were significantly associated with smartphone medical app use.

Conclusions: Most respondents have a smartphone medical app and have used it in clinical practice. Attitude, internet access, computer training, perceived usefulness of the app, information technology support staff, and technical skill are the most notable factors that are associated with smartphone medical app use by physicians.

(JMIR Mhealth Uhealth 2021;9(3):e19310) doi:10.2196/19310

KEYWORDS

application; medical; physician; smartphone; mobile phone

Introduction

The main sources of information for health care professionals at the point of care were once textbooks [1]. However, health care professionals increasingly use smartphone medical apps for patient care, clinical reference, and education [2]. The smartphone is a tool that has recently grown in use and has been accepted by health professionals and medical students. It is a new technology that has an operating system, the capability of installing various apps, and the ability to do complex calculations and establish related communications at the point of care [1,3]. Smartphone apps are tools that can be downloaded onto smartphones or computer tablets and enhance patient care, increase efficiency, or provide individualized learning for clinicians [4]. In 2011, Apple created the Apps for Healthcare Professionals section within the medical category of the iTunes App Store, a unique feature among mobile app marketplaces [5]. Smartphones have a wide range of uses from the internet to email; they offer on-the-go access to information that was never before possible [6]. The most commonly used smartphone medical apps are UpToDate and Medscape, as shown in Table 1.

Table 1. Summary of the most commonly used smartphone medical apps.

Rank	App	App category	Operating system
1	UpToDate	Medical reference	iOS, Android, Windows
2	Medscape	Medical reference	iOS, Android, Windows
3	Epocrates	Drug and medical reference	iOS, Android, Windows
4	PEPID	Decision support/reference	iOS, Android, Windows
5	Figure 1	Medical image	iOS, Android, Windows
6	MedCalc	Drug reference	iOS, Android, Windows
7	Prognosis	Decision support/reference	iOS, Android, Windows
8	Skyscape	Drug and medical reference	iOS, Android, Windows
9	Diseases Dictionary Medical	Medical disorders & diseases with detailed definitions, symptoms, causes, and treatment information	iOS, Android, Windows
10	Calculate by QxMD	Literature and drug reference	iOS, Android, Windows

A study conducted in the United Kingdom showed that smartphone medical apps like British National Formulary, eLogbook, and medical calculator (MedCalc) have been commonly used by physicians [7]. This technology can lead to improved decision making, reduced numbers of medical errors, and improved communication between hospital medical staff [8-10].

Another study from the United Kingdom stated that due to the ease of use of smartphone medical apps, 18.5% of doctors made suggestions to their colleagues to use apps as a quick reference during clinical practice [7]. However, lack of support and updating of apps by their developers, lack of adequate skill to use apps, lack of creating motivation in using apps, and problems related to security and confidentiality of patient information have undermined the use of smartphone medical apps at the point of care [10-12]. A study conducted in Korea showed low use of smartphone medical apps by physicians [13].

Evidence shows that medical app use is high in high-income countries; compared with the Korean study, a study conducted in the United Kingdom found higher medical app use (72.4%) among doctors [14].

Another study in the United States reported that 56% of physicians use apps in their clinical practice. There was a decreasing trend in app use with increased training level, and the most useful app types included textbook and reference materials (average response: 55%), classification and treatment algorithms (46%), and general medical knowledge (43%); there was a greater desire for apps among residents than among fellows and attending physicians [15]. This might be due to residents being less experienced physicians than the senior specialists and subspecialists; as a result, they need some assistance from colleagues and seniors, so medical apps are immediate reference tools that can be accessed anywhere during clinical practice. This explains why the level of use among residents is higher than among their senior attending physicians.

A study conducted in Saudi Arabia showed the use patterns of smartphone medical apps among residents at clinical practice for counselling and clinical communication (50%), among interns for drug reference (56%), and among externs for resources and e-books (65%) [16].On the other hand, a study done in Iran reported that the most popular medical apps were Medscape and UpToDate, and 61.3% of the physicians were using their apps more than once a day, mostly for drug information [17].

According to a cross-sectional study done in Ghana, over 43.1% of physicians frequently used medical apps on their smartphones for clinical decision making, which shows relatively low use of medical apps compared with that in high-income countries [18].

Due to different factors, physicians remain reluctant to adopt these technologies in clinical practice [2]; the most common factors that affect the use of smartphone medical apps are behavioral factors (information technology (IT)–related experience, attitude, computer-related skill) [19], factors related to medical app characteristics (perceived usefulness, perceived ease of use, privacy and security concerns), organizational factors (infrastructure, IT support, and computer-related training) [20]. Shreds of evidence revealed that underutilization of apps in clinical practice by health care professionals is due to a lack of technical skill [21]. According to the findings of a cross-sectional study conducted in the United Kingdom, appearing to be looking at a phone during clinical practice could be misinterpreted as checking emails or using social networks by colleagues and patients [7,22-25].

Security and privacy are the key factors of the functionality of any mHealth system [26]; unfortunately, most of the time these important areas are neglected by development teams of mHealth systems, and the majority of currently available mHealth apps impart little or no security. This will affect the use of these apps significantly [27-29]. Perceived ease of use is another factor that determines the use of smartphone medical apps; those with a user-friendly interface are more likely to be used [23]. The aim of this study was to assess the level of smartphone medical app use and associated factors among physicians in referral hospitals of the Amhara region, Ethiopia.

Methods

Study Design and Setting

This was a cross-sectional, questionnaire-based study done to assess smartphone medical app use and associated factors among physicians in referral hospitals (Gondar University, Felege Hiwot, Debre Markos, Dessie, and Debre Birhan referral hospitals) in Amhara region, Ethiopia, which consists of 10 administrative zones, 1 special zone, 181 woredas, and 78 urban centers. The capital city of the state of Amhara is Bahir-Dar [30]. It is located in the northwestern and north-central parts of Ethiopia. The state shares common borders with the state of Tigray in the north, Afar in the east, Oromia in the south, Benishangul/Gumuz in the southwest, and the Republic of Sudan in the west [31].

The sample size was computed as 423, which was 80% of the total population during the data collection period, using a single population proportion formula taking 50% at a 95% confidence level and assuming a 5% margin of error and 10% nonresponse rate. There are 5 referral hospitals in the region; all 5 referral hospitals were included, and then proportional allocation was made for each hospital. Finally, a simple random sampling method was used. 213 permanent doctors (general practitioners and specialists), 95 residents, and 115 interns were recruited using a lottery method (each member of the population was assigned a unique number, each number was written on a separate piece of paper or card of the same size, the cards were mixed well in a basket, and the sample was drawn) and formed the sample.

The survey consisted of 36 questions encompassing the following domains: (1) sociodemographic characteristics, (2) attitude, (3) factors related to medical app characteristics (perceived usefulness and perceived ease of use of apps), (4) physician technical skill, and (5) organizational factors [20].

Attitude was assessed using a Likert scale (1=strongly disagree, 2=disagree, 3=neutral, 4=agree, 5=strongly agree) for the following questions: Do you believe smartphone medical apps give you greater control over your work schedule? Do you believe smartphone medical apps will improve your work performance? Do you think smartphone medical apps allow you to conduct your job more quickly?

Organizational factors were assessed with yes/no questions as follows: Do you have internet access in your office at clinical practice? Have you ever taken any training on use of smartphone medical apps? Does your organization have IT support staff?

Perceived usefulness and ease of use of medical apps were assessed using a Likert scale (1=strongly disagree, 2=disagree, 3=neutral, 4=agree, 5=strongly agree) for questions such as the following: Do you think learning to use smartphone medical apps will not require much time? Do you think smartphone medical apps are easy to use? Do you think smart phone medical apps are easy to use during your consultations with patients?

Smartphone medical app use was assessed using 9 categories of medical apps: (1) drug reference, (2) clinical score systems, (3) disease diagnosis, (4) procedure documentation, (5) literature search, (6) clinical communication, (7) health information system clients, (8) medical training, and (9) web browsing. Physicians who used 5 or more of the app categories, which was the median, were categorized as having used smartphone medical apps and those who used fewer than 5 were not.

A self-administered questionnaire was adopted from previous studies [14,32]. To ensure the validity of the questionnaire, an expert panel (10 doctors having at least 5 years' experience in general practice or primary care research) was invited to review the tool and revise it, and reliability was calculated to be α=.78.

Before the actual data collection, pilot testing of the questionnaire was conducted among 20 physicians at Debre Tabor hospital to check internal consistency within the questioners. Then necessary correction was done based on the pretest finding.

Two days' training was given for 5 data collectors on the objective of the study and data collection procedures. Data was collected from January 15 to March 30, 2019, using self-administered questionnaires; one data collector was assigned for each hospital, and the supervisor facilitated the data collection process. The principal investigator and supervisors did daily supportive supervision on data collectors. Data backup activities, such as storing data at different places and putting data in different formats (hard and soft copies), were performed to prevent data loss.

Study Variables and Operational Definitions

Dependent Variable

The dependent variable was the physician's smartphone medical app use.

Independent Variables

The independent variables were sociodemographic factors (age, sex, profession, educational status, experience), medical app–related factors (perceived usefulness, perceived ease of use, privacy and security concerns), organizational factors (internet access, IT support, computer-related training), and behavioral factors (knowledge, attitude, technical skill, IT-related experience).

Operational Definitions

In this study, "physician" includes general practitioners, residents, dentists, specialists, and subspecialists.

Smartphone is a class of mobile phone with multipurpose mobile computing capability and features like high-definition camera, third-party app installation, and global positioning system [33].

Medical apps are computer programs or software apps that are designed to run on a mobile device such as a smartphone or tablet and are meant to be used for clinical purposes.

Smartphone medical apps are medical apps designed to run specifically on smartphones [34].

Study participants who scored at or above the median of 5 out of the 9 categories of Food and Drug Administration–approved medical apps were categorized as having used smartphone medical apps [18].

For the attitude questions, study participants who scored the median or higher on the 5-point Likert scale were categorized as having a good attitude, and those who scored below the median were categorized as having a poor attitude [35].

For the perceived usefulness questions, study participants who scored the median or higher on the 5-point Likert scale were categorized as thinking smartphone medical apps were useful for their job, and those who scored below the median were categorized as thinking smartphone medical apps were not useful for their job [36].

For the perceived ease of use questions, study participants who scored the median or higher on the 5-point Likert scale were categorized as thinking smartphone medical apps were easy to use, and those who scored below the median were categorized as thinking smartphone medical apps were not easy to use.

Data Processing and Analysis

Data were entered into Epi Info, version 7 (Centers for Disease Control and Prevention), and exported to SPSS, version 21 (IBM Corp), for further analysis. Descriptive statistics were computed to summarize variables, and the binary logistic regression model was used to measure the association between dependent and independent variables. Both crude odds ratios for binary logistic regression analysis and adjusted odds ratios (AOR) for multivariable logistic regression analysis were estimated with 95% CIs to show the strength of associations. Finally, a *P* value of less than .05 in the multivariable logistic regression analysis was used to identify variables significantly associated with the use of smartphone medical apps.

Ethical Considerations

In conducting the study, ethical clearance was obtained from the University of Gondar ethical review board. Additional permissions to access participants were also obtained from each hospital administrator. In addition, written informed consent was gained from all participants (Multimedia Appendix 1), participation in the study was voluntary, and no incentive was provided for the participants.

Results

Sociodemographic Characteristics

A total of 417 physicians were included in this study with a response rate of 98.6% (417/423). Two-thirds (275, 65.9%) of the respondents were male. The mean age was 33 years (SD 8) with the majority in the age group of 25-34 years. More than three-fourths (375, 89.9%) of the physicians had the medical app installed on their smartphones (Table 2).

Table 2. Sociodemographic characteristics of respondents working at referral hospitals of Amhara region, North Ethiopia, 2019 (N=417).

Variable	n (%)
Gender	
Male	275 (65.9)
Female	142 (34.1)
Age (years)	
≤30	217 (52.0)
>30	200 (48.0)
Educational status	
General practitioner	219 (52.5)
Resident	127 (30.5)
Specialist	71 (17.0)
Department	
Internal medicine	80 (19.2)
Pediatrics	54 (12.9)
Radiology	28 (6.7)
Surgery	67 (16.1)
Ophthalmology	25 (6.0)
Gynecologist	65 (15.6)
Dermatology	18 (4.3)
ENT[a]	36 (8.6)
Other	44 (10.6)
Work experience (years)	
1-3	231 (55.4)
3-6	91 (21.8)
>6	95 (22.8)
Medical app ownership	
Yes	375 (89.9)
No	42 (10.1)

[a]ENT: ear, nose, and throat.

Smartphone Medical App Use of Physicians at Referral Hospitals

According to this study, 63.3% (264) of the respondents reported that they use apps in their clinical practice (95% CI 58.3%-67.9%), and the most commonly used smartphone medical app category was diagnosis/management (62%) (Table 3).

According to this study, the most commonly used smartphone app was UpToDate (300/417, 71.9%) (Table 4).

Most study respondents (354, 85%) used their apps daily, while 10.5% used them 3 times a week (Table 5).

Table 3. Smartphone medical app use at referral hospitals among physicians, 2019 (N=417).

Medical app types	Education level			
	General practitioner, n	Resident, n	Specialist, n	Total, n (%)
Disease diagnosis	134	84	43	261 (62.6%)
Literature search	120	67	31	218 (52.3%)
Browsing	113	54	29	196 (47.0%)
HIS[a] clients	106	54	35	195 (46.8%)
Clinical score system	104	71	34	209 (50.1%)
Medical training	104	59	31	194 (46.5)
Drug reference	97	48	28	173 (41.5)
Clinical communication	87	58	23	168 (40.3%)
Procedure documentation	73	51	28	152 (36.5%)

[a]HIS: health information system.

Table 4. Medical apps used by physicians working at referral hospitals in Amhara regional state, Ethiopia, 2019.

App used	n (%)
UpToDate	
Yes	305 (73.1)
No	112 (26.9)
Medscape	
Yes	276 (66.2)
No	141 (33.8)
MedCalc	
Yes	226 (54.2)
No	192 (46.0)
Doximity	
Yes	104 (24.9)
No	313 (75.1)
PEPID	
Yes	50 (12.0)
No	367 (88.0)
Case	
Yes	97 (23.3)
No	320 (76.7)
Figure 1	
Yes	99 (23.7)
No	318 (76.3)
Read by QxMD	
Yes	90 (21.6)
No	327 (78.4)

Table 5. Frequency of smartphone medical app use among physicians working at referral hospitals of Amhara regional state, 2019 (N=375).

Medical app	Frequency of use, n (%)			
	Daily	Three times a week	Once a week	I don't know
UpToDate	333 (88.8)	28 (7.5)	12 (3.2)	2 (0.5)
Medscape	335 (89.3)	25 (6.7)	13 (3.5)	2 (0.5)
MedCalc	346 (92.3)	18 (4.8)	9 (2.4)	2 (0.5)
Doximity	340 (90.7)	23 (6.1)	12 (3.2)	0
Figure 1	339 (90.4)	24 (6.4)	12 (3.2)	0
Read by QxMD	335 (89.3)	27 (7.2)	13 (3.5)	0
Case	346 (92.3)	21 (5.6)	8 (2.1)	0
PEPID	322 (85.9)	18 (4.8)	35 (9.3)	0

Factors Associated With Smartphone Medical App Use Among Physicians

A total of 6 variables were selected as potential predictors of smartphone app use after bivariable logistic regression and entered multivariable logistic regression. Included variables were attitude, internet access, computer training, past IT experience, perceived ease of use of the app, perceived usefulness of the app, the technical skill of the physicians, and availability of IT support staff, which were positively related to smartphone medical app use by physicians at referral hospitals in Amhara region.

In this study, physicians with a favorable attitude toward smartphone medical apps were 1.64 times more likely to use them than physicians with an unfavorable attitude were (AOR 1.64, 95% CI 1.05-2.55). Similarly, physicians who have IT support staff at hospitals were 2.36 times more likely to be smartphone medical app users compared to their counterparts (AOR 2.36, 95% CI 1.5-3.08) (Table 6).

Table 6. Bivariable and multivariable regression analysis of factors with smartphone medical app use among physicians in referral hospitals of Amhara regional state, North Ethiopia, 2019 (N=417).

Variable	App use, n (%)		Crude OR[a] (95% CI)	AOR[b] (95% CI)	*P* value
	Yes	No			
Education level					
GP[c]	142 (34.1)	77 (18.5)	1.51 (0.87-2.60)	1.68 (0.91-3.10)	—[d]
Resident	83 (19.9)	44 (10.6)	1.54 (0.85-2.80)	1.70 (0.88-3.30)	—
Specialist	39 (9.4)	32 (7.7)	1	1	—
Internet access					
Yes	210 (50.4)	88 (21.1)	2.87 (1.85-4.45)	2.82 (1.75-4.50)	<.001
No	54 (12.9)	65 (15.6)	1	1	—
Computer training					
Yes	179 (42.9)	72 (17.3)	2.36 (1.57-3.56)	1.71 (1.09-2.67)	<.001
No	85 (20.4)	81 (19.4)	1	1	—
IT support staff					
Yes	156 (37.4)	58 (13.9)	2.366 (1.57-3.56)	2.363 (1.50-3.08)	.001
No	108 (25.9)	95 (22.8)	1	1	—
Technical skill					
Yes	225 (54)	103 (24.7)	2.8 (1.73-4.52)	2.54 (1.50-4.30)	<.001
No	39 (9.4)	50 (12.0)	1	1	—
Attitude					
Yes	153 (36.7)	69 (16.5)	1.67 (1.12-2.50)	1.64 (1.05-2.55)	.01
No	111 (26.6)	84 (20.1)	1	1	—
Perceived usefulness					
Yes	158 (37.9)	70 (16.8)	1.76 (1.18-2.64)	1.65 (1.06-2.56)	.02
No	106 (25.4)	83 (19.9)	1	1	—

[a]OR: odds ratio.

[b]AOR: adjusted odds ratio.

[c]GP: general practitioner.

[d]Not available.

Discussion

Principal Findings

This study assessed the use of smartphone medical apps and associated factors among physicians at referral hospitals in the Amhara region. Out of 417 participants, 375 (89.9%) have a medical app installed on their mobile device, and disease diagnosis/management was the most commonly used medical app category. Factors like attitude, perceived usefulness, internet access, and past computer-related training were found to be associated with smartphone medical app use. In this study, the use rate of smartphone medical apps by physicians was 63.3% (95% CI 58.3%-67.9%). This result was consistent with that of a study done in Britain (60%) [6]. On the other hand, it was lower than those of a study done in Canada (77.0%) [11], a study conducted in Germany at the Leipzig Medical School (68%) [37], and a study conducted in the United States on the American Society of Plastic Surgeons (72%) [38]. This might be due to well-organized infrastructure at the clinical practice site, awareness of physicians on the use of smartphone medical apps for patient care, and availability of technological guidelines that promote the use of smartphone medical apps in America, Canada, and Germany. On the other hand, the result is higher than that of a study in Ghana (43.1%) [18]. A possible explanation may be due to the sample size difference (the sample size of the study in Ghana was 65) and the study period (this study was conducted about 4 years ago).

This study indicated that the medical apps most frequently used by physicians were UpToDate, Medscape, and MedCalc. This result is in line with the study conducted in Canada [11]. Most of the physicians (271/375, 72.3%) preferred smartphone medical apps as an information source for reference during clinical practice. The proportion is lower in the study in Ghana than in this study; this might be due to the accessibility of smartphones in our setup and sample size difference (the sample size of the study in Ghana was 65) [18].

This study found that the perceived usefulness of smartphone medical apps was positively associated with medical app use among physicians at referral hospitals in the Amhara region (P=.04). This is in line with a study conducted in Taiwan [39]. Perceived usefulness of apps was a significant determinant of app use according to a study conducted in a Malaysian public hospital [40]. That result is comparable with this current study (OR 1.65, 95% CI 1.06-2.56]).

This study revealed that physicians who had good technical skills were 2.54 times more likely to use smartphone medical apps at clinical practice than those who had poor technical skills (AOR 2.54, 95% CI 1.50-4.30]). A study conducted in Czech Republic also indicated that technical skill was a factor for smartphone app use [41]. This might be because people with good technical skills are more receptive to new technology and capable of operating new apps.

According to this study, physicians who were working in an institution with internet access (WiFi) were 2.82 times more likely to use smartphone medical apps than those who had no internet access (AOR 2.82, 95% CI 1.75-4.54). This might be because the availability of internet access makes the regular update of medical apps easier and makes it possible to exchange information through a medical app, such as for consultation among senior physicians.

Physicians who were in an institution that has IT support staff were 2.36 times more likely to use smartphone medical apps than their counterparts (AOR 2.363, 95% CI 1.50-3.08).

This study revealed that the odds of physicians with favorable attitudes being users of smartphone medical apps were 1.64 times higher than those of their counterparts (AOR 1.64, 95% CI 1.05-2.55), which is in line with the result obtained by a study conducted in Iran [35].

This implies that the attitude of physicians is key in the implementation of such apps in clinical practice. From the results above, we found that smartphone medical apps used by physicians did well in providing relevant medical information during clinical practice and received positive reviews from physicians. However, in other aspects (ie, outside of improving clinical decision making, saving time, helping to make differential diagnoses, performing useful medical-related calculations, and providing faster access to evidence-based medical practices or cases), medical apps did not meet the needs of physicians well, as most of the medical apps are not freely accessible, the cost of these apps is not affordable, and more importantly, the payment mechanism is not available in our country, Ethiopia. Therefore, in the future, there will be much room for improvement, and health care institutions in resource-limited countries like Ethiopia should offer an institutional access mechanism to such medical apps that is accessible freely.

Conclusion

The findings of this study showed that smartphone medical app use was 63.3%. Favorable attitude, internet access, computer training, perceived usefulness of the app, the technical skill of the physicians, and availability of IT support staff were the most notable factors that were associated with smartphone medical app use.

Based on this result, smartphone medical apps have inevitable contributions to successful and effective clinical practice. To effectively use this technology in clinical practice, health care organizations should create awareness of its use and implications in health care service, improve internet connectivity, and provide training on the use of these apps.

Acknowledgments

We would like to thank the University of Gondar Institute of Public Health for the approval of the ethical clearance and the referral hospitals and directors for giving us permission to collect data. We also forward gratitude to the data collectors, supervisors, and study participants.

This work would not be possible without the financial support of the Doris Duke Charitable Foundation under grant number 2017187. The mission of the Doris Duke Charitable Foundation is to improve the quality of people's lives through grants supporting the performing arts, environmental conservation, medical research, and child well-being and through preservation of the cultural and environmental legacy of Doris Duke's properties.

Authors' Contributions

GHT conceived of the study and coordinated data collection. GHT, BCT, ATA, and HAG performed statistical analysis and drafted the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Informed consent statement.
[DOCX File , 16 KB - mhealth_v9i3e19310_app1.docx]

References

1. Mosa ASM, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012;12:67 [FREE Full text] [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]
2. Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. P T 2014 May;39(5):356-364 [FREE Full text] [Medline: 24883008]
3. Wu RC, Morra D, Quan S, Lai S, Zanjani S, Abrams H, et al. The use of smartphones for clinical communication on internal medicine wards. J Hosp Med 2010;5(9):553-559. [doi: 10.1002/jhm.775] [Medline: 20690190]
4. Rodrigues M, Visvanathan A, Murchison J, Brady R. Radiology smartphone applications; current provision and cautions. Insights Imaging 2013 Oct;4(5):555-562 [FREE Full text] [doi: 10.1007/s13244-013-0274-4] [Medline: 23912880]
5. Thomairy NA, Mummaneni M, Alsalamah S, Moussa N, Coustasse A. Use of Smartphones in Hospitals. Health Care Manag (Frederick) 2015;34(4):297-307. [doi: 10.1097/HCM.0000000000000080] [Medline: 26506291]
6. Lindquist AM, Johansson PE, Petersson GI, Saveman B, Nilsson GC. The use of the Personal Digital Assistant (PDA) among personnel and students in health care: a review. J Med Internet Res 2008;10(4):e31 [FREE Full text] [doi: 10.2196/jmir.1038] [Medline: 18957381]
7. Patel RK, Sayers AE, Patrick NL, Hughes K, Armitage J, Hunter IA. A UK perspective on smartphone use amongst doctors within the surgical profession. Ann Med Surg (Lond) 2015 Jun;4(2):107-112 [FREE Full text] [doi: 10.1016/j.amsu.2015.03.004] [Medline: 25905017]
8. Prgomet M, Georgiou A, Westbrook JI. The impact of mobile handheld technology on hospital physicians' work practices and patient care: a systematic review. J Am Med Inform Assoc 2009;16(6):792-801 [FREE Full text] [doi: 10.1197/jamia.M3215] [Medline: 19717793]
9. Wu R, Rossos P, Quan S, Reeves S, Lo V, Wong B, et al. An evaluation of the use of smartphones to communicate between clinicians: a mixed-methods study. J Med Internet Res 2011;13(3):e59 [FREE Full text] [doi: 10.2196/jmir.1655] [Medline: 21875849]
10. Armstrong DG, Giovinco N, Mills JL, Rogers LC. FaceTime for Physicians: Using Real Time Mobile Phone-Based Videoconferencing to Augment Diagnosis and Care in Telemedicine. Eplasty 2011 May 03;11:e23 [FREE Full text] [Medline: 21559249]
11. Boruff JT, Storie D. Mobile devices in medicine: a survey of how medical students, residents, and faculty use smartphones and other mobile devices to find information. J Med Libr Assoc 2014 Jan;102(1):22-30 [FREE Full text] [doi: 10.3163/1536-5050.102.1.006] [Medline: 24415916]
12. Jebraeily M, Fazlollahi Z, Rahimi B. The Most Common Smartphone Applications Used By Medical Students and Barriers of Using Them. Acta Inform Med 2017 Dec;25(4):232-235 [FREE Full text] [doi: 10.5455/aim.2017.25.232-235] [Medline: 29284911]
13. Safdari R, Jebraeily M, Rahimi B, Doulani A. Smartphone medical applications use in the clinical training of medical students of UMSU and its influencing factors. Eur J Exp Biol 2014;4(1):633-637.
14. Payne KFB, Wharrad H, Watts K. Smartphone and medical related App use among medical students and junior doctors in the United Kingdom (UK): a regional survey. BMC Med Inform Decis Mak 2012 Oct 30;12:121 [FREE Full text] [doi: 10.1186/1472-6947-12-121] [Medline: 23110712]
15. Franko OI, Tirrell TF. Smartphone app use among medical providers in ACGME training programs. J Med Syst 2012 Oct;36(5):3135-3139. [doi: 10.1007/s10916-011-9798-7] [Medline: 22052129]
16. Alfawareh HM, Jusoh S. Smartphones usage among university students: Najran University case. IJAR 2014 Mar 30;6(2):321-326. [doi: 10.7813/2075-4124.2014/6-2/B.48]
17. Jahanshir A, Karimialavijeh E, Sheikh H, Vahedi M, Momeni M. Smartphones and Medical Applications in the Emergency Department Daily Practice. Emerg (Tehran) 2017 Jan 09;5(1):e14 [FREE Full text] [Medline: 28286821]
18. Senya KY. Use of Smart Phone Applications for Clinical Decision Making Among Junior Medical Practitioners in Selected Health Facilities in The Greater Accra Region, Ghana. Ghana: University of Ghana; 2015 Nov 11. URL: http://ugspace.ug.edu.gh/bitstream/handle/123456789/8208/KAFUI%20YAYRA%20SENYA_USE%20OF%20SMART%20PHONE%20APPLICATIONS%20FOR%20CLINICAL_2015.pdf [accessed 2019-06-07]
19. Jeon E, Park H. Factors affecting acceptance of smartphone application for management of obesity. Healthc Inform Res 2015 Apr;21(2):74-82 [FREE Full text] [doi: 10.4258/hir.2015.21.2.74] [Medline: 25995959]
20. Gagnon M, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):212-220 [FREE Full text] [doi: 10.1093/jamia/ocv052] [Medline: 26078410]
21. Zahra F, Hussain A, Mohd H. Factors affecting mobile health application for chronic diseases. J Telecommun Electron Comput Eng 2018;10:81.
22. Chen J, Park Y, Putzer G. An examination of the components that increase acceptance of smartphones among healthcare professionals. Electron J Heal Informatics 2010;5(2):1-12. [doi: 10.4018/978-1-4666-9828-4.les10]
23. Lee S, Choi MJ, Rho MJ, Kim D, Choi IY. Factors Affecting User Acceptance in Overuse of Smartphones in Mobile Health Services: An Empirical Study Testing a Modified Integrated Model in South Korea. Front Psychiatry 2018;9:658 [FREE Full text] [doi: 10.3389/fpsyt.2018.00658] [Medline: 30631283]

24. Shenouda JEA, Davies BS, Haq I. The role of the smartphone in the transition from medical student to foundation trainee: a qualitative interview and focus group study. BMC Med Educ 2018 Jul 31;18(1):175 [FREE Full text] [doi: 10.1186/s12909-018-1279-y] [Medline: 30064424]
25. Nematollahi M, Faghiri K, Barati O, Bastani P. Smart Phone Acceptance among Physicians: Application of Structural Equation Modelling in the Largest Iranian University. J Biomed Phys Eng 2017 Mar;7(1):87-94 [FREE Full text] [Medline: 28451582]
26. Kameda-Smith MM, Iorio-Morin C, Winkler-Schwartz A, Ahmed US, Bergeron D, Bigder M, Canadian Neurosurgery Research Collaborative (CNRC). Smartphone Usage Patterns by Canadian Neurosurgery Residents: A National Cross-Sectional Survey. World Neurosurg 2018 Mar;111:e465-e470. [doi: 10.1016/j.wneu.2017.12.089] [Medline: 29277596]
27. BinDhim NF, Trevena L. Health-related smartphone apps: regulations, safety, privacy and quality. BMJ Innov 2015 Mar 05;1(2):43-45. [doi: 10.1136/bmjinnov-2014-000019]
28. Chanpimol S, Seamon B, Hernandez H, Harris-Love M, Blackman M. Using Xbox kinect motion capture technology to improve clinical rehabilitation outcomes for balance and cardiovascular health in an individual with chronic TBI. Arch Physiother 2017;7 [FREE Full text] [doi: 10.1186/s40945-017-0033-9] [Medline: 28824816]
29. Zahra F, Hussain A, Mohd H. Usability factors of mobile health application for chronic diseases. 2016 Presented at: The International Conference on Applied Science and Technology 2016; April 11-13, 2016; Kedah, Malaysia.
30. HSFR/HFG Project. HSFR/HFG End of Project Achievement Highlights - Amhara.: Health Finance and Governance Project, Abt Associates; 2018 Jun. URL: https://www.slideshare.net/HFGProject/hsfrhfg-end-of-project-regional-report-amhara [accessed 2019-06-01]
31. Amhara National Regional State Bureau of Health. For Supply, Training and Installation of Solar photovoltaic Systems For Rural Health posts in North Gondar Zone. Ethiopian Herald. Bahir Dar: Ethiopian Press Agency; 2009 Jul 04. URL: http://www.ethiomarket.com/bid/2009/July09/bid_jul_4_amhara_national_regional_state_bureau_of_health.htm [accessed 2019-06-07]
32. Ma Q, Liu L. The Technology Acceptance Model: A Meta-Analysis of Empirical Findings. In: Advanced Topics in End User Computing, Volume 4. Hershey, PA: IGI Global; 2005:112-128.
33. Putzer GJ, Park Y. Are physicians likely to adopt emerging mobile technologies? Attitudes and innovation factors affecting smartphone use in the Southeastern United States. Perspect Health Inf Manag 2012;9:1b [FREE Full text] [Medline: 22737094]
34. Talwar YK, Karthikeyan S, Bindra N, Medhi B. Smartphone - A User-friendly Device to Deliver Affordable Healthcare - A Practical Paradigm. J Health Med Inform 2016;7(3) [FREE Full text] [doi: 10.4172/2157-7420.1000232]
35. Nematollahi M, Faghiri K, Barati O, Bastani P. Smart Phone Acceptance among Physicians: Application of Structural Equation Modelling in the Largest Iranian University. J Biomed Phys Eng 2017 Mar;7(1):87-94 [FREE Full text] [Medline: 28451582]
36. Yasini M, Marchand G. Mobile Health Applications, in the Absence of an Authentic Regulation, Does the Usability Score Correlate with a Better Medical Reliability? Stud Health Technol Inform 2015;216:127-131. [Medline: 26262024]
37. Sandholzer M, Deutsch T, Frese T, Winter A. Predictors of students' self-reported adoption of a smartphone application for medical education in general practice. BMC Med Educ 2015 May 21;15:91 [FREE Full text] [doi: 10.1186/s12909-015-0377-3] [Medline: 25994310]
38. Grow JN, Vargo JD, Nazir N, Korentager R. Smartphone Applications in Plastic Surgery: A Cross-Sectional Survey of 577 Plastic Surgeons, Fellows, Residents, and Medical Students. Aesthet Surg J 2019 Nov 13;39(12):NP530-NP537. [doi: 10.1093/asj/sjz052] [Medline: 30810739]
39. Hsiao J, Chen R. Understanding Determinants of Health Care Professionals' Perspectives on Mobile Health Continuance and Performance. JMIR Med Inform 2019 Mar 18;7(1):e12350 [FREE Full text] [doi: 10.2196/12350] [Medline: 30882353]
40. Ganasegeran K, Renganathan P, Rashid A, Al-Dubai SAR. The m-Health revolution: Exploring perceived benefits of WhatsApp use in clinical practice. Int J Med Inform 2017 Jan;97:145-151. [doi: 10.1016/j.ijmedinf.2016.10.013] [Medline: 27919374]
41. Elavsky S, Smahel D, Machackova H. Who are mobile app users from healthy lifestyle websites? Analysis of patterns of app use and user characteristics. Transl Behav Med 2017 Dec;7(4):891-901 [FREE Full text] [doi: 10.1007/s13142-017-0525-x] [Medline: 28929368]

Abbreviations

AOR: adjusted odds ratio
IT: information technology
mHealth: mobile health

Edited by L Buis; submitted 13.04.20; peer-reviewed by M Hassen, K Gashu, K Ganasegeran; comments to author 12.06.20; revised version received 26.10.20; accepted 22.01.21; published 26.03.21.

Please cite as:
Teferi GH, Tilahun BC, Guadie HA, Amare AT
Smartphone Medical App Use and Associated Factors Among Physicians at Referral Hospitals in Amhara Region, North Ethiopia, in 2019: Cross-sectional Study
JMIR Mhealth Uhealth 2021;9(3):e19310
URL: https://mhealth.jmir.org/2021/3/e19310
doi:10.2196/19310
PMID:33769303

Original Paper

User-Centered Design Process of an mHealth App for Health Professionals: Case Study

Amarasinghe Arachchige Don Nalin Samandika Saparamadu[1*], MBBS, DipIBLM; Piyum Fernando[2*], BSc; Peizi Zeng[3], BSc; Henry Teo[3], BSc, MHA; Andrew Goh[3], MSc; Joanne Mee Yin Lee[3], BSc, MSc; Choong Weng Leslie Lam[3], MBBS, FRCPath, FAACC, DABCC

[1]MOH Holdings Private Limited, Singapore, Singapore

[2]School of Arts, Media, and Engineering, Arizona State University, Tempe, AZ, United States

[3]Department of Laboratory Medicine, Ng Teng Fong General Hospital, Singapore, Singapore

[*]these authors contributed equally

Corresponding Author:
Amarasinghe Arachchige Don Nalin Samandika Saparamadu, MBBS, DipIBLM
MOH Holdings Private Limited
1 Maritime Square
Singapore, 099253
Singapore
Phone: 65 84980484
Email: samandika@icloud.com

Abstract

Background: User-centered design processes are infrequently employed and not fully explored for building mobile health (mHealth) apps that are particularly targeted to health professionals as end users. The authors have used a user-centered design–based approach to build an mHealth app for health professionals, tasked to deliver medical laboratory-related information on a daily basis.

Objective: Our objective is to generate a simple and functional user-centered design process for mHealth apps for health professionals. This paper presents the key learnings from design activities.

Methods: A stratified random sample of doctors and nurses was recruited for the study. The design activities were planned in the following sequence: focus group discussion for situation analysis and information architecture, design activity 1 for wireframe designing, design activity 2 for wireframe testing, and user testing sessions 1 and 2.

Results: The final design and functions of the app, information architecture, and interactive elements were largely influenced by the participatory design–based user-centered design activities. As a result of the design process, we could identify the mental models of processing requests for information and personal preferences based on the experience. These findings were directly or indirectly incorporated into the app design. Furthermore, finding alternative ways of working within time constraints and cultural barriers and the methods employed to manage the challenges of interdisciplinary discourse stood out among the lessons learned.

Conclusions: We recommend a user-centered design process based on a participatory design approach in mHealth app design, enriched with focus group discussions where possible.

(JMIR Mhealth Uhealth 2021;9(3):e18079) doi:10.2196/18079

KEYWORDS

user-centered design; participatory design; mobile health applications; mHealth; smartphones; health professionals; healthcare; human-computer interaction; mobile phones

Introduction

Background

The term mobile health (mHealth) is defined as emerging mobile communications and network technologies for health care systems, which is an exponentially growing market in the smartphone era [1,2]. Current estimates suggest that there are more than 40,000 mHealth apps [3]. The global mHealth market is projected to grow at a compound annual growth rate of around 35.65% over the next few years to reach approximately US $115.61 billion by 2025 [4].

Many current mHealth interventions for health care–related issues are designed on the basis of existing healthcare system constructs, and they may not be as effective as those that involve end users in the design process [5]. In recent times, user-centered design–based approaches have been reportedly used for building mHealth apps, most of which are focused on chronic diseases [6-9], cancer [10,11], health and well-being, lifestyle interventions [12-14], mental health [15], sexual health [16], pain management [10,17], and remote patient monitoring [18]. Commonly used user-centered design frameworks for building mHealth apps include Information Systems Research [19], the Health Information Technology Usability Evaluation Model [20], and the BUS (behavior change theories, user-centered design, and social marketing) framework [21].

Despite the growth of user-centered design–based mHealth apps built for patients and the general public, there has been little exploration of how to use them for health professionals as end users. Hence, little is known about the unique challenges and opportunities for applying user-centered design to build apps for health professionals.

Our findings are drawn upon a real-world design process where the authors have employed a participatory design–based user-centered design approach to build an mHealth app targeting health professionals. Traditionally, laboratory-related information is made available to the staff through a laboratory service manual, which is a 217-page document accessed via the hospital intranet, describing over 525 laboratory investigations and procedures available at the hospital. The idea of delivering this information through an mHealth app was conceived as a result of an audit conducted to describe the phone call patterns received at the laboratory and a follow-up survey conducted among health professionals that revealed the towering demand for laboratory-related information and the infrequent use of the laboratory service manual by end users who were inclined toward an mHealth intervention [22]. Combined, these findings set out the primary objective of this app, which is to replace the existing laboratory service manual and to create an efficient, convenient platform to deliver medical laboratory-related information to health professionals at Ng Teng Fong General Hospital, Singapore, on a daily basis.

Health professionals' frequent dilemmas related to laboratory tests in clinical practice take many forms: deciding the adequacy of the size of a specimen according to laboratory's rejection criteria, clarifying the correct method of collection and specimen container, decisions related to the urgency of investigations and results, etc. In these scenarios, laboratory-related information such as specimen type, minimum sample size, appropriate specimen container or tube, method of collection, special instructions (eg, fasting blood sample), turnaround time, price, etc, play an important role in decision making.

The contribution of this paper is twofold: first, we describe our methods, materials, results, and outcomes. Then, by reflecting on our own experiences, we discuss the insights and implications of our work through the challenges faced, lessons learned, and strategies adopted to overcome challenges.

Objectives

Our objective is to generate a simple and functional user-centered design process for mHealth apps for health professionals (specifically, for an app that delivers medical laboratory-related information).

Methods

Overview

To achieve our objective, we aimed to employ (a) an intermediate approach to user-centered design and participatory design by direct and indirect involvement of our end users (doctors and nurses) at various stages of the design process and (b) multiple user-centered design methods and to tailor the whole process according to the unique needs of our end users. In this section, we describe the methods in sampling, data analysis, and designing activities and sessions, as well as the design procedures.

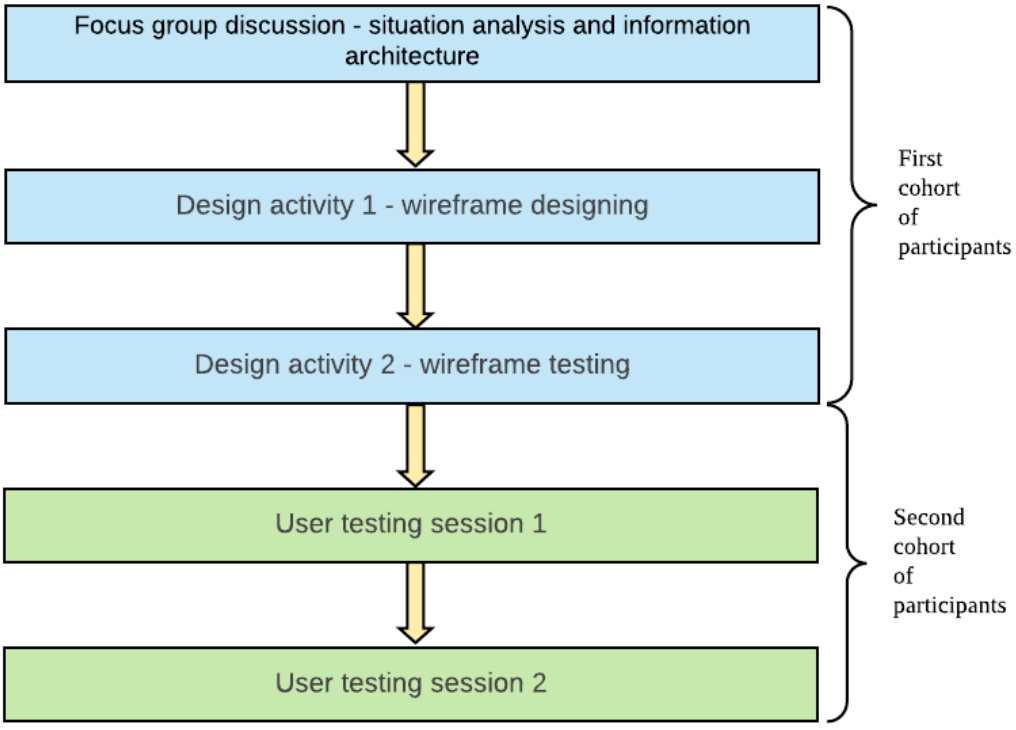
A stratified random sampling model was employed by the study team—the project team and a human-computer interaction (HCI) consultant—to select participants for the sessions. Strata were defined to include inpatient settings (medical, surgical wards, and the ambulatory unit) and outpatient settings or clinics. We recruited doctors and nurses in rough proportion to their workforce ratios. However, the emergency department was excluded in view of significant time constraints and perceived manpower shortage (Table 1). Informed written consent was obtained from all participants throughout the process prior to each session or design activity (Figure 1).

Table 1. Demographic data of participants.

Characteristic	First cohort (focus group discussion and design activities)	Second cohort (user testing sessions)	Total
Participants, n	15	9	24
Age (years), mean (SD)	35.4 (8.1)	34.8 (7.6)	__[a]
Gender, n (%)			
Female	12 (80)	6 (67)	18 (75)
Male	3 (20)	3 (33)	6 (25)
Nationality, n (%)			
Singaporeans	9 (60)	2 (22)	11 (46)
Non-Singaporeans	6 (40)	7 (78)	13 (54)
Ethnicity, n (%)			
Chinese	7 (47)	3 (33)	10 (42)
Malay	3 (20)	1 (11)	4 (17)
Indian	3 (20)	1 (11)	4 (17)
Others	2 (13)	4 (44)	6 (25)
Role in hospital, n (%)			
Physician	5 (33)	2 (22)	7 (29)
Nurse	10 (67)	7 (78)	17 (71)
Setting, n (%)			
Inpatient/ward	9 (60)	6 (67)	15 (62)
Outpatient clinic	4 (27)	0 (0)	4 (17)
Ambulatory clinic	2 (13)	3 (33)	5 (21)

[a]Not available.

Figure 1. An illustration of the sequence of design activities and sessions.



Statistical analysis of demographic data was carried out on IBM SPSS, version 23.0 (IBM Corp). Categorical variables were expressed as frequencies and percentages, and continuous variables were expressed using means and standard deviation.

Following each session, all study data, including audio recordings and handwritten notes, were transcribed verbatim without participant identifiers and annotated with pauses and nonverbal expressions. Anonymized transcripts were analyzed by the study team to create and examine codes, themes, and categories. Transcripts were coded independently by two individuals using line-by-line coding technique. The codes were compared, and the differences were discussed among them. Affinity mapping technique was employed for thematic analysis and establishing categories, and the findings, outcomes, and insights were generated based on this analysis.

To further ensure reliability of the study, effective session facilitating techniques were utilized: well-prepared facilitators (social and conversational skills) and rapport building with the participants.

Focus Group Discussion: Situation Analysis and Information Architecture

The objective of the first session was to understand the existing user behavior and practices related to laboratory information usage (situation analysis) and obtain initial user inputs to determine what laboratory information should be presented and how it should be structured and highlighted in the app (information architecture). In addition, we used this session as a means of building rapport and goodwill between participants and the facilitators.

This session was designed to be conducted in a span of an hour. First, we randomly divided all the participants into three subgroups by counting numbers; the same method has been used in the activities mentioned throughout the paper. In the first activity, the participants were given a scenario of a less frequently used laboratory investigation requested by a physician during a busy morning ward round. We asked them to write down the first things that came to their minds on sticky notes while they were fulfilling this task. We requested them to produce at least five responses per group. Following this activity, one group member from each group was asked to share the responses on the whiteboard.

In the next activity, the same groups were asked to draw a flowchart to explain the process of dispatching a blood sample as part of their routine daily work. Groups were given an example of how to perform this task, and they were encouraged to simulate a real-life scenario of their choice. Members were asked to describe their flowcharts to other groups and were given the opportunity to ask questions.

The third activity was designed to prioritize the information available in the laboratory service manual to be included in the app based on user input. We provided details of 10 uncommon laboratory tests available in the laboratory service manual and asked participants to use color codes (red, yellow, and green) to highlight information based on following criteria: green for details that they already knew from memory, red for details that they did not know unless they referred to the manual, and yellow for details for which they needed reassurance or confirmation. We carefully handpicked 10 tests, considering the scarcity of information on the type of specimen, tube or container, use of preservatives or transport media, dispatch instructions, other special instructions, etc.

Finally, we prioritized the information based on the user responses using a four-quadrant matrix under two main domains (how frequently it is used and how well it is known to the user) in order to summarize them (ie, mostly known vs least known and least used vs mostly used).

Design Activity 1: Wireframe Designing

The first design activity was planned in the form of a co-design activity. The same cohort of participants was present in the second session, except for 1 participant who had called in sick. She was replaced by another member from the same ward.

This session was an hour-long design activity. It consisted of three main activities focused on determining the information architecture and designing the initial wireframes (skeletal visual representations of the app screens). Participants were randomly divided into two groups, and each group was given a set of selected laboratory investigations written on A5-size papers. It had a balanced representation of the mainstream disciplines of laboratory medicine: chemical pathology, microbiology, anatomic pathology, hematology, immunology, genetics, and molecular diagnostics. The set of tests given to each group were not identical but comparable. They were asked to classify all the tests handed to them within 10 minutes. Following this, they were asked to share the principles of the classification that they employed, and each group was given a chance to ask questions of the other group. Both groups were given another chance to reclassify investigations in a different way and in a shorter period of time (5 minutes).

Following this activity, we asked participants to jot down at least five different keywords that they would use to search for each investigation on the reverse side of each paper. This was aimed at gathering user inputs toward designing the search function of the app.

In the third activity, both groups were asked to sketch wireframes depicting how they envisioned the app's home page, visual representation of test details, and search functionality. An in-depth discussion was carried out to understand the reasons behind each wireframe design.

Design Activity 2: Wireframe Testing

The same cohort of participants from the previous design activity attended the third session, with the addition of 2 new participants. Data from the previous session were analyzed and discussed with the study team. Based on the outcomes of the previous session, two sets of wireframe mockups were created by the HCI consultant: one with a home page containing the categories of tests (categorized by discipline and by specimen type) and a directly accessible list of frequently used tests and the other with a home page with links to go to “categories” and “all tests” and a list of frequently used tests.

Once again, the participants were randomly divided into two groups. An open discussion on design, ease of use, order and

presentation of information, user-friendliness, and specific functions was conducted among groups with open-ended questions (Table 2). We also asked some specific questions that the project team and HCI specialist had during mockup wireframe preparations to shed some light on uncertainties.

Table 2. Questionnaire for design activity 2, wireframe testing.

Serial number	Script
Home page	
1	Please go to the home page [hand them the first prototype] and tell me how you would feel if you got to use this design/user interface on a smartphone app?
2	What do you think about this home page design?
3	Think of/give some example tests that you would like to add to this list. How often do you think you would access a test using this list? Where do you like to have the list of bookmarked tests in your mobile app? Why do/don't you think it is useful to have this list on the home page?
4	Will an "all tests" function be useful? Alternatively, would you prefer to access "all tests" using the search icon below? Why?
5	How do you want to see new notifications on your app? How important are these notifications to you?
6	[Now, the interviewer hands them the second prototype of the home page.] What do you think about this design/user interface?
7	What do you think about this design compared to the previous one?
8	Do you think that you need an icon for "most common tests" that would by default have some commonly used tests across all disciplines listed in it and would be customizable [you can add or remove items from this list]? Alternatively, do you want something like "my tests," which would come empty with the app but be customizable according to your needs?
Categories	
1	These categories were carefully selected depending on your suggestions and discussions from the last week. What are your thoughts about this categorization? [Repeat this question for each categorization.]
2	Do you have any other suggestions for categorizations?
3	Can you take a few minutes and think of any test that would not easily fit into any of these categories? [Please explain that the idea of the categories is fluid and one test could appear in more than one category if necessary, overlapping.]
4	Any other comments/suggestions?
Search function	
1	As per our discussions from the last session, we realized that the search function should be robust. Here are the shortlisted "keywords" in order of importance that are suggested to be used in the smart search engine: medical domain/discipline and related diseases. [Interviewer gives a real-life scenario.] What do you think about this search order? What are the changes/modifications you would like to introduce to this search function?
2	Do you want to see your old search items and history before you start your search? Why?
3	Any other comments/suggestions?
Test-details page	
1	What do you think about the arrangement of information under each test [ask about spacing, font size, graphics, pictures of the tube, etc]?
2	Any comments on the order of information? Any suggestions to change this interface?
3	Is there anything else that you would like to see on this page [information or graphical presentation]?
4	What do you think about the picture of the tube? Any other details that you would like to see in this image?
5	How do you want to add this test into your bookmarked list? What is your preferred workflow?
6	This design uses a "tick" on the right upper corner. What kind of icon (or user interface element) would you prefer?
Hamburger menu	
1	[A wireframe of a hamburger menu is to be handed to each group.] Could you try and list the things that you want to see/ would like to see on this page?
2	[Once participants have finished listing, facilitate a discussion on each listed item.] Why do you want this item on the hamburger menu, and what are the reasons why it's best suited in the menu page but not elsewhere (ie, home page)?

User Testing Session 1

Within 5 months from the third session, the first functional prototype of the app was developed based on the user inputs gathered since inception. The first user testing session was conducted to evaluate the usability of the first functional prototype. An all-new group of nurses and doctors were recruited using stratified random sampling methods. Out of 10 participants, 5 used iOS and 5 used Android operating systems. One Android user dropped out of the study due to privacy concerns related to installing an unpublished external app on the phone. This incident did not have an impact on the final design of the app as the process of publishing spontaneously resolves the concern.

All the participants were given a copy of the prototype laboratory mobile app downloaded on their mobile phones. They were asked to perform a set of specific tasks in the shortest possible time, and the time taken by each participant was recorded. A discussion was carried out with each group using open-ended questions. Furthermore, at the end of the session, we obtained feedback from the participants on general topics (Table 3).

Table 3. Questionnaire for user testing session 1.

Serial number	Script	Allocated time (min)
Verbal and visual instructions given to all participants		13-15
1	Find out “special instructions” for taking a blood culture.	
2	Assume that you find a piece of false information on a “test page.” Submit feedback regarding this issue using the app. [E.g., a test page of full blood count says that the sample has to be sent in a “yellow tube.”]	
3	Find the contact information to arrange a “frozen section.”	
4	Search and find a test to help diagnose Wilson disease. [Hint: You can probably find this test on the app even if you do not remember the name of the test.]	
5	Go to categories “by discipline,” create a new category, and give that new category a name. Then, add three (3) tests of your choice into the new category that you created.	
6	Go to the category that you created and remove one (1) test from the category.	
7	Find the “order of draw” on the app.	
8	Go and find the types of tubes/containers that you can use to send specimens to check “acetaminophen level” of a patient.	
9	Check how many phone notifications you have received on your app.	
Questions to be asked of participants after each task is performed		25
1	Tell us about your experience of performing this task.	
2	What are the difficulties you faced while performing this task?	
3	I like [the things that I like about this feature/function].	
4	I wish [the things that I wish to see in this feature/function].	
5	Ask more questions in detail depending on the scenario (eg, What are your concerns regarding the current design? How would you suggest improving it? What would you like to see instead of the current design/workflow?).	
General thoughts to be asked at the end of the session		5-10
1	Any further suggestions to improve the functionality of the app?	
2	What do you think about the information provided on the app? Do you see anything missing? Do you find anything that is not useful?	
3	How do you like the look and feel of the app, such as the colors, fonts, and user interface elements like buttons and menus? Any suggestions to improve them?	
4	Are there any other thoughts/suggestions/comments you would like to add that have not been discussed so far?	

User Testing Session 2

Within 3 months from the first user testing session, the second prototype of the app was prepared. The same cohort of participants who took part in user testing session 1 were invited to take part in this activity. A total of 9 participants took part in this session.

Participants were asked to download a copy of the latest version of the laboratory mobile app, to use it at work, and to try to replace their daily work process with the app. After allowing them to use the app for a week, we interviewed them individually using an open-ended questionnaire.

Results

Overview

The mean ages of our participants were 35.4 and 34.8 years for the group that participated in focus group discussions and design activities and for the group that participated in user testing sessions, respectively. The youngest member of the whole cohort was a 24-year-old medical officer, while the oldest member was a 53-year-old assistant nurse clinician. Singapore being a multiethnic, multicultural city-state, we observed a representation of all three major ethnic groups (ie, Chinese, Malay, Indian) as well as a fair representation of migrant workers who contribute to Singapore health care from the following countries of origin: China, India, the Philippines, Malaysia, South Korea, and Sri Lanka (Table 1). Moreover, all members of the groups were smartphone users, and all of them had more than one year of experience with the electronic medical records of the hospital, Epic System (Epic Systems Corporation).

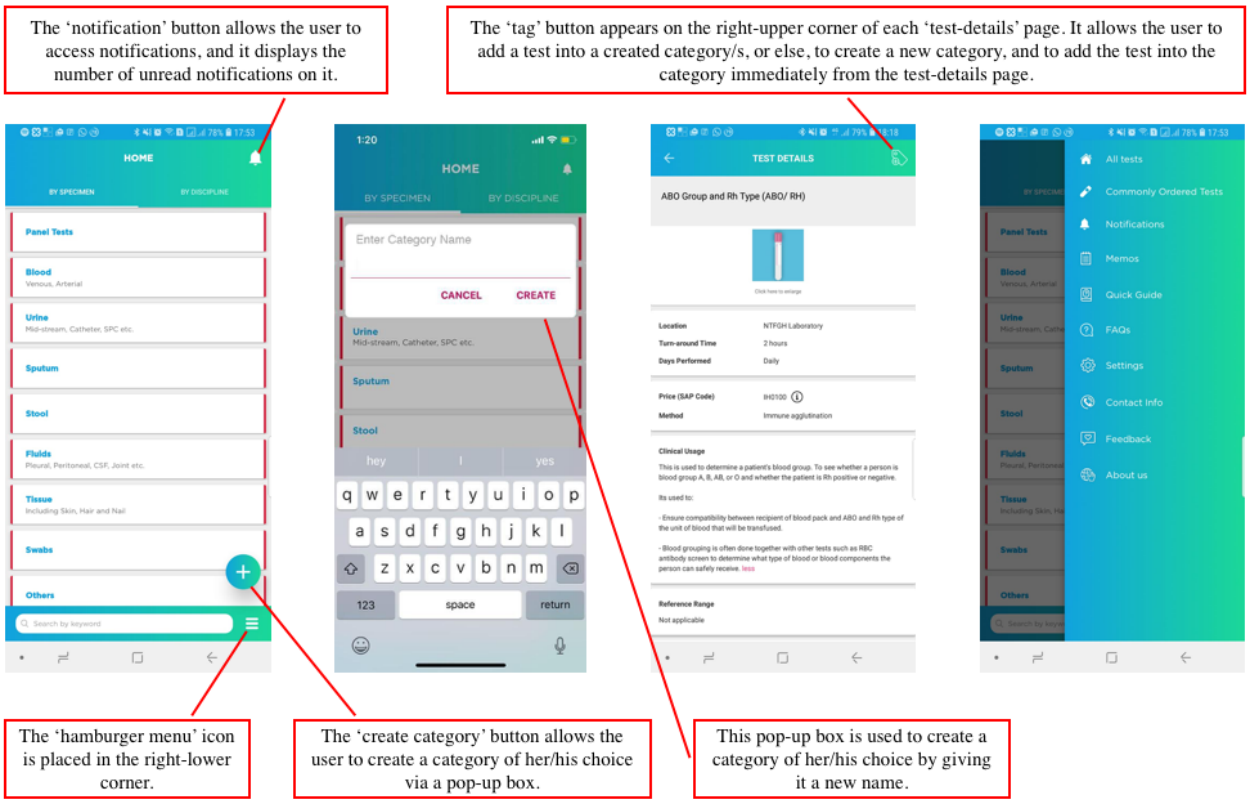
Focus Group Discussion: Situation Analysis and Information Architecture

From this session, we identified existing behaviors and habits of our users. We also observed significant differences in the use of terminology and work processes between nurses and doctors, which in turn generated valuable insights into ensuring a greater level of customizability of the app. In addition, we identified that the levels of importance related to variables of laboratory tests (ie, test details) are understood by all subgroups in similar patterns, making certain pieces of information salient while making the remainder less important. This information was directly used in determining the visual layouts of the 'test-details' page.

Design Activity 1: Wireframe Designing

These activities helped to foreground two commonly used criteria to classify laboratory tests: by specimen and by discipline. These outcomes were directly incorporated into the design of the home page (Figure 2). Furthermore, nurses preferred to search for investigations by the specimen whereas doctors preferred to search by the related discipline; however, both groups preferred to have their own lists of investigations to refer to instantly, as well as a smart search engine.

Figure 2. First version of the laboratory mobile app after focus group discussions and design sessions.



Surprisingly, the wireframe designs of the "home page" and "test-details" page were strikingly similar between the two groups. They both suggested to include categories of tests and a search button on the home page, assuming that users will mostly use the search button instead of looking for tests in categories, and a starred or quick access list. Both groups prioritized details such as special instructions, pictures of relevant test tubes, containers, and preservatives, and the price of the tests, in that order. One group suggested that it would be useful to display a list of related tests under each test, and the other group was in agreement.

Participatory wireframe design activities helped the study team to further identify the features that can be useful for users. For example, a periodically updated list of commonly ordered tests was eventually added to the app based on the ideas generated from wireframe designing sessions, as it can be useful for new staff as reference material.

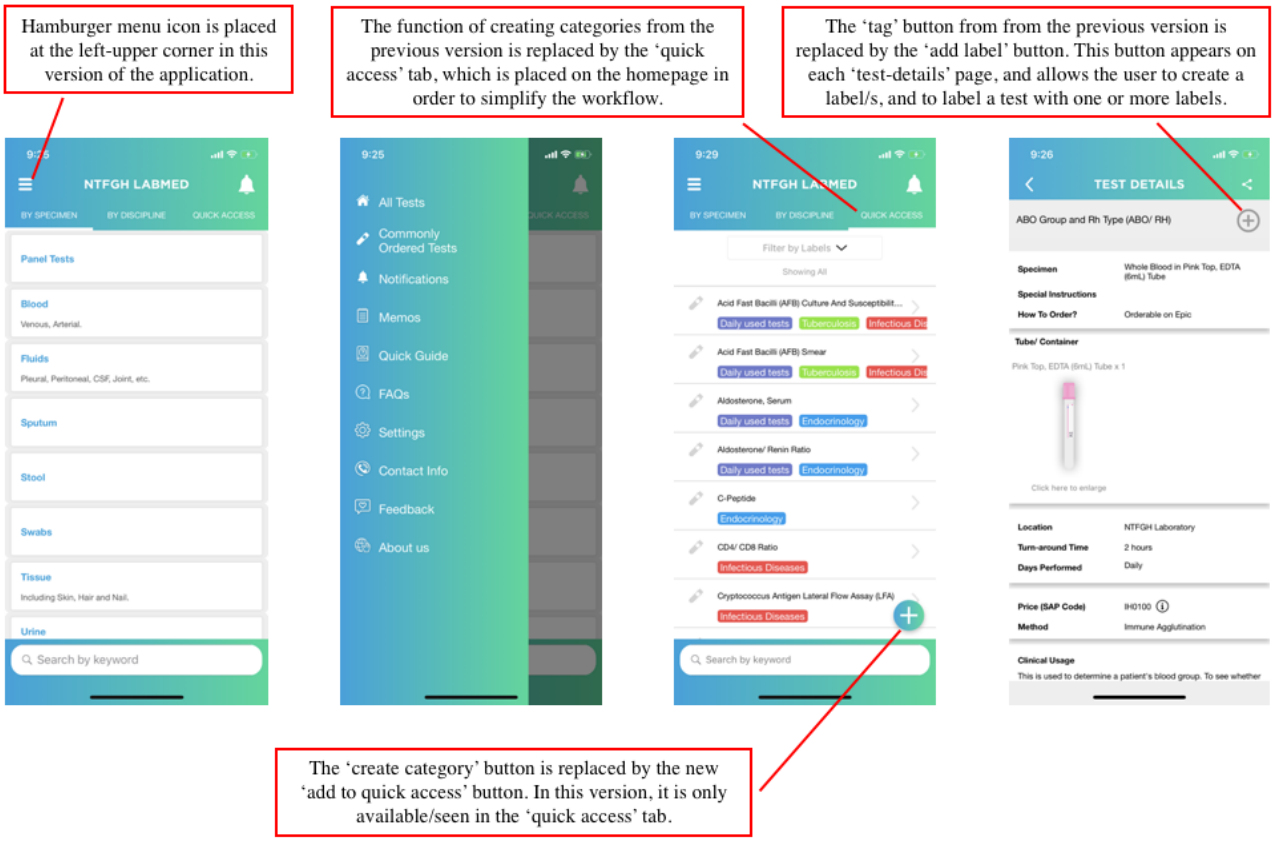
Design Activity 2: Wireframe Testing

Wireframe testing triggered the participants to express their preferences and comment on potential improvements that can be made to the wireframe designs presented to them. For example, participants preferred to see a home page with all the test categories displayed on it, as opposed to the alternative wireframe. In addition, reference ranges were added, and a photograph of the tube or container was suggested over an illustration with regard to the "test-details" page. Outcomes of the third session are shown in Figure 2.

User Testing Session 1

We identified the specific pain points for users when they perform certain tasks with the mobile app through this session. For example, users pointed out the process of creating categories as a cumbersome workflow and suggested changes. In addition, they suggested moving the hamburger menu icon that was placed on the right-lower corner to the left-upper corner, as it is commonly observed in popular local e-commerce apps. Outcomes of the fourth session are shown in Figure 3.

Figure 3. Second version of the laboratory mobile app after user testing sessions.



User Testing Session 2

The second user testing session revealed that the users were generally satisfied with the functions and user interface of the app, and there were no specific suggestions to ameliorate the design (Figure 3). Hence, it was decided by the project team that the app was ready for launch.

Discussion

General Takeaways

The final design of the app, including the features, functions, information architecture, interactive elements, and aesthetics, were largely influenced by the participatory design–based user-centered design activities. As a result of the design process, we could identify the existing mental models of the end users related to laboratory tests, level of significance of laboratory-related information, how requests for information are dealt with, and how additional information is requested via telephone calls. There were personal preferences depending on the knowledge base, experience, setting, and job scope. These findings were directly or indirectly incorporated in the final design of the app. Furthermore, it triggered the need for in-depth customization while maintaining generally used mental models in the user interface of the app.

In parallel to the outcomes of the user-centered design activities, the final design of the app was influenced by the inputs from the HCI specialist based on universal user experience design principles. These inputs were incorporated in implementing several key features of the app. For example, simplified illustrations of specimen containers (eg, test tubes) were used in place of actual photographs to avoid information overload and to emphasize distinct features of each container type. Similarly, a quick access one-touch search function and the ability to assign custom labels to tests were also added based on the HCI specialist's suggestions.

By reflecting on our user-centered design process and our own experience of working with multiple stakeholders including the project team, an HCI consultant, health professionals, and an

external developer team, we discuss key takeaways of this case study under four categories below.

Cultural Considerations

Despite the hierarchy that is observed in the medical culture [23], we observed more collaborative work and interdependency during design activities and discussions. However, it was observed that mostly doctors started leading the group during the technical discussions including information architecture, and nurses were eager to get the doctors involved. During wireframe testing and participatory design activities, the nurses were more inclined to pen down their ideas, and doctors were more collaborative despite leading the group. Disagreements between group members were dealt with through ongoing discussions; however, in certain instances we observed outlying ideas being rejected by doctors and experienced nurses without their merits being carefully examined.

These behavioral patterns were prominent during the first session, but during design sessions, they were noted to have eventually diminished. Possible reasons could be enhanced group dynamics over the course of the sessions and the nature of activities during different sessions. For example, the first focus group discussion was more dependent on technical knowledge and experience, whereas medical or technical knowledge did not have a bearing on the design activities.

Designing the Study Around Constraints

While attempting to make focus group discussions, design sessions, and user testing sessions equally accessible and ensure fair participation, we faced major challenges with time constraints, making it the most significant challenge to mention [24,25].

In response to these legitimate limitations, the study team conducted a mock focus group discussion with elective nursing students who were rotated to the Department of Laboratory Medicine as part of their polytechnic training. Furthermore, facilitators trained themselves with each other while giving running commentary and feedback. These activities were carried out to see how well the participants could understand quick instructions, as well as to evaluate timeliness of activities and discussions. Mock focus group discussions helped us to fine-tune the construction and wording of questions and instructions to better relate with our participants culturally [26,27] and to make them succinct to save time.

We also carried out a quick feedback activity after each focus group discussion and design session to better understand the concerns of the participants. As a result of the feedback, snacks were arranged for subsequent sessions, which further encouraged participation. It was also conveyed by the participants that it's important to commence and conclude sessions right on time in view of serious time constraints that they were facing [24,25].

Benefits of Participatory Design Approach for Designing Information Architecture

A well-designed information architecture helps users find information efficiently and complete tasks easily (ie, finding information on a particular blood test by navigating through the app user interface). In our case, information that is intended to be presented in the proposed mHealth app (ie, laboratory test information) has already been stored in end users' memories to various degrees as they refer to that information during daily hospital routines. Therefore, the key goal of co-designing the information architecture with end users was to understand their existing mental models of this information and try to build upon them. The outcomes of the co-design activity, where the end users categorized and labeled a selected set of laboratory tests, provided valuable insights to this end.

While there are many instances where co-designing approaches have been successfully utilized for designing information architecture [28,29], our experience sheds light on two specific benefits of such approaches for mHealth apps. First, co-design activities similar to ours can help user experience designers to become familiar with medical procedures and terminology without spending too much time on secondary research. For example, in our case, the user input gathered through co-design activities remarkably eliminated the need for designers to understand the content of the laboratory service manual. Second, such activities can help designers thoroughly understand the existing user behaviors related to information usage and align the final information architecture of the mHealth app to be consistent with them. In our case, insights gained from co-design activities ultimately allowed us to design an information architecture that resonated with the one that users were accustomed to, which in turn significantly lowered the barriers for adopting the new mobile app into their daily routines.

Role of the Facilitator in the Design Process: Facilitating Communication and Collaboration Among Multiple Stakeholders

Due to highly domain-specific content and medical terminology related to the laboratory tests, the developer team and HCI consultant had to go through a steep learning curve to familiarize themselves with the context of the mHealth app. The hierarchy, organizational structure, and bureaucracy of the hospital environment also made this process somewhat challenging for these external stakeholders. To this end, having a medical professional with experience in participatory and qualitative research as a facilitator significantly helped the external parties to overcome those challenges and understand the context quickly. Provision of an overview of the organizational structure, culture, and existing workflows, coordinating with internal stakeholders, co-conducting activities, and working around cultural barriers are some examples of the roles of the facilitator.

Conclusions

We highly recommend a user-centered design process, based on a participatory design approach to designing mHealth apps for health professionals. Focus group discussions will be an added advantage in the design process, provided opportunity is materialized. A broad understanding of the culture, hierarchy, and bureaucracy is pivotal in planning participatory design activities; however, working on team dynamics through ice breaking and building rapport is a cornerstone of success.

In addition, a test run for polishing and fine-tuning the process is certainly a strength. Furthermore, the role of a facilitator with a medical and qualitative research background could most likely

add value as well as convenience in situations where there is a steep learning curve and a race against time.

Acknowledgments

The researchers acknowledge the tremendous contributions toward focus group discussions, design activities, and user testing sessions by the nurses and doctors (participants). We acknowledge the microbiology, clinical chemistry, hematology, and anatomic pathology sections of the laboratory for their most valuable support in content creation. We also recognize the enormous support rendered by Mrs Lynn Low, secretary of the Department of Laboratory Medicine, with personal communications and administrative tasks and the support extended by Ms Kimberly Price (linguistics specialist) with proofreading the manuscript.

This work was funded by the JurongHealth Internal Research & Development Grant Award FY2016 and FY2017 (project code number: 16-55).

Conflicts of Interest

None declared.

References

1. Istepanian RSH, Lacal JC. Emerging mobile communication technologies for healthome imperative notes on m-health, in: Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2003;2?6. 2003 Sep 17 Presented at: 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 17-21 Sept. 2003; Cancun, Mexico p. 1414.
2. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
3. Aitken M, Gauntlett C. Patient apps for improved healthcare: from novelty to mainstream?. Parsippany, NJ: IMS Institute for Healthcare Informatics; Apr 09, 2014:30-30.
4. Market Data Forecast. Global mHealth market / Mobile Health Market Size, Share, Trends, Growth Analysis Report - Segmented By Type, Services, Application & Region – Industry Forecast (2020 to 2025). Hyderabad, India: Market Data Forecast; Feb 2020.
5. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;Suppl:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
6. LeRouge C, Wickramasinghe N. A review of user-centered design for diabetes-related consumer health informatics technologies. J Diabetes Sci Technol 2013 Jul 01;7(4):1039-1056 [FREE Full text] [doi: 10.1177/193229681300700429] [Medline: 23911188]
7. Cafazzo JA, Casselman M, Hamming N, Katzman DK, Palmert MR. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res 2012 May 08;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]
8. Hilliard ME, Hahn A, Ridge AK, Eakin MN, Riekert KA. User Preferences and Design Recommendations for an mHealth App to Promote Cystic Fibrosis Self-Management. JMIR Mhealth Uhealth 2014 Oct 24;2(4):e44 [FREE Full text] [doi: 10.2196/mhealth.3599] [Medline: 25344616]
9. Cafazzo JA, Leonard K, Easty AC, Rossos PG, Chan CT. The user-centered approach in the development of a complex hospital-at-home intervention. Stud Health Technol Inform 2009;143:328-333. [Medline: 19380956]
10. Mirkovic J, Kaufman DR, Ruland CM. Supporting cancer patients in illness management: usability evaluation of a mobile app. JMIR Mhealth Uhealth 2014 Aug 13;2(3):e33 [FREE Full text] [doi: 10.2196/mhealth.3359] [Medline: 25119490]
11. Griffin L, Lee D, Jaisle A, Carek P, George T, Laber E, et al. Creating an mHealth App for Colorectal Cancer Screening: User-Centered Design Approach. JMIR Hum Factors 2019 May 08;6(2):e12700 [FREE Full text] [doi: 10.2196/12700] [Medline: 31066688]
12. van der Weegen S, Verwey R, Spreeuwenberg M, Tange H, van der Weijden T, de Witte L. The development of a mobile monitoring and feedback tool to stimulate physical activity of people with a chronic disease in primary care: a user-centered design. JMIR Mhealth Uhealth 2013 Jul 02;1(2):e8 [FREE Full text] [doi: 10.2196/mhealth.2526] [Medline: 25099556]
13. Curtis KE, Lahiri S, Brown KE. Targeting Parents for Childhood Weight Management: Development of a Theory-Driven and User-Centered Healthy Eating App. JMIR Mhealth Uhealth 2015 Jun 18;3(2):e69 [FREE Full text] [doi: 10.2196/mhealth.3857] [Medline: 26088692]
14. Gabrielli S, Dianti M, Maimone R, Betta M, Filippi L, Ghezzi M, et al. Design of a Mobile App for Nutrition Education (TreC-LifeStyle) and Formative Evaluation With Families of Overweight Children. JMIR Mhealth Uhealth 2017 Apr 13;5(4):e48 [FREE Full text] [doi: 10.2196/mhealth.7080] [Medline: 28408361]
15. Dekker MR, Williams AD. The Use of User-Centered Participatory Design in Serious Games for Anxiety and Depression. Games for Health Journal 2017 Dec;6(6):327-333. [doi: 10.1089/g4h.2017.0058]

16. Chipambwa W, Terera K. Designing a User Centered M-Health Application Interface to be Used for HIV/AIDS Management in Zimbabwe. 2016 Nov 01 Presented at: 7th International Conference on Appropriate Technology; 2016; Victoria Falls (Mosi Oa Tunya), Zimbabwe.
17. Stinson JN, Jibb LA, Nguyen C, Nathan PC, Maloney AM, Dupuis LL, et al. Development and testing of a multidimensional iPhone pain assessment application for adolescents with cancer. J Med Internet Res 2013 Mar 08;15(3):e51 [FREE Full text] [doi: 10.2196/jmir.2350] [Medline: 23475457]
18. Smaradottir B, Gerdes M. User-Centered Design of a COPD Remote Monitoring Application; Experiences from the EU-project United4Health. International Journal on Advances in Software 2015;8:350-360.
19. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
20. Brown W, Yen P, Rojas M, Schnall R. Assessment of the Health IT Usability Evaluation Model (Health-ITUEM) for evaluating mobile health (mHealth) technology. J Biomed Inform 2013 Dec;46(6):1080-1087 [FREE Full text] [doi: 10.1016/j.jbi.2013.08.001] [Medline: 23973872]
21. Patel S, Arya M. The BUS Framework: A comprehensive tool in creating an mHealth App utilizing Behavior Change Theories, User-Centered Design, and Social Marketing. J Mob Technol Med 2017 Apr;6(1):39-45 [FREE Full text] [doi: 10.7309/jmtm.6.1.6] [Medline: 29104674]
22. Lam L, Saparamadu AADNS, Zeng P. An Audit of Telephone Calls Received in a Singapore Public Hospital Clinical Laboratory. 2019 Presented at: 71st AACC Annual Scientific Meeting and Clinical Lab Expo of the American Association for Clinical Chemistry; August 4-8, 2019; Anaheim, CA.
23. Helmreich R, Merritt A. Culture at Work in Aviation and Medicine: National, Organizational, and Professional Influences. Culture at Work in Aviation and Medicine: National, Organizational, and Professional Influences. National, Organizational and Professional Influences. 2nd edition. Routledge; 2001.
24. Lim J, Bogossian F, Ahern K. Stress and coping in Singaporean nurses: a literature review. Nurs Health Sci 2010 Jun;12(2):251-258. [doi: 10.1111/j.1442-2018.2010.00514.x] [Medline: 20602699]
25. Yang S, Meredith P, Khan A. Stress and burnout among healthcare professionals working in a mental health setting in Singapore. Asian J Psychiatr 2015 Jun;15:15-20. [doi: 10.1016/j.ajp.2015.04.005] [Medline: 25922279]
26. Paulston C, Kiesling S, Rangel E. The Handbook of Intercultural Discourse and Communication. In: Blackwell handbooks in linguistics. Malden, MA: Wiley-Blackwell; 2012.
27. Carson JG, Nelson GL. Writing groups: Cross-cultural issues. Journal of Second Language Writing 1994 Jan;3(1):17-30. [doi: 10.1016/1060-3743(94)90003-5]
28. Baek J, Lee K. A participatory design approach to information architecture design for children. CoDesign 2008 Sep;4(3):173-191. [doi: 10.1080/15710880802281026]
29. Massanari AL. Designing for imaginary friends: information architecture, personas and the politics of user-centered design. New Media & Society 2010 Feb 09;12(3):401-416. [doi: 10.1177/1461444809346722]

Abbreviations

HCI: human-computer interaction
mHealth: mobile health

Edited by L Buis; submitted 02.02.20; peer-reviewed by L Ratnayake, J D'Arcey, B Chaudhry; comments to author 29.03.20; revised version received 19.06.20; accepted 10.08.20; published 26.03.21.

Please cite as:
Saparamadu AADNS, Fernando P, Zeng P, Teo H, Goh A, Lee JMY, Lam CWL
User-Centered Design Process of an mHealth App for Health Professionals: Case Study
JMIR Mhealth Uhealth 2021;9(3):e18079
URL: https://mhealth.jmir.org/2021/3/e18079
doi:10.2196/18079
PMID:33769297

Tutorial

Using a Clinical Workflow Analysis to Enhance eHealth Implementation Planning: Tutorial and Case Study

Stephanie Staras[1,2], MSPH, PhD; Justin S Tauscher[3], MSc, PhD; Natalie Rich[1,4], MD; Esaa Samarah[1,2,5], MSW; Lindsay A Thompson[1,2,6], MD, MS; Michelle M Vinson[7], MSc; Michael J Muszynski[7], MD; Elizabeth A Shenkman[1,2], PhD

[1]Department of Health Outcomes and Biomedical Informatics, College of Medicine, University of Florida, Gainesville, FL, United States

[2]Institute for Child Health Policy, University of Florida, Gainesville, FL, United States

[3]Behavioral Research in Technology and Engineering Center, Department of Psychiatry, University of Washington, Seattle, WA, United States

[4]Department of Radiology, College of Medicine, University of Florida, Gainesville, FL, United States

[5]College of Social Work, Florida State University, Tallahassee, FL, United States

[6]Department of Pediatrics, College of Medicine, University of Florida, Gainesville, FL, United States

[7]Department of Clinical Sciences, Florida State University College of Medicine Orlando Regional Campus, Orlando, FL, United States

Corresponding Author:
Stephanie Staras, MSPH, PhD
Department of Health Outcomes and Biomedical Informatics
College of Medicine
University of Florida
2004 Mowry Rd Room 2238
Gainesville, FL
United States
Phone: 1 3522948299
Email: sstaras@ufl.edu

Abstract

eHealth apps often fail to improve clinical outcomes due to poor integration with clinical workflow—the sequence and personnel needed to undertake a series of tasks for clinical care. Our central thesis is that eHealth interventions will be more effective if the clinical workflow is studied and taken into consideration for intervention implementation. This paper aims to provide an introductory tutorial on when and how to use a clinical workflow analysis to guide the implementation of eHealth interventions. The tutorial includes a step-by-step guide to conducting a clinical workflow analysis in planning for eHealth implementation. We began with a description of why a clinical workflow analysis is best completed before the implementation of eHealth interventions. Next, we described 4 steps needed to perform the clinical workflow analysis: the identification of discrete workflow components, workflow assessment, triangulation, and the stakeholder proposal of intervention implementation. Finally, we presented a case study of a clinical workflow analysis, which was conducted during patient visits of patients aged 11 or 12 years from 4 diverse pediatric or family medicine clinics to plan the implementation of a tablet-based app for adolescent vaccination. Investigators planning the implementation of new eHealth interventions in health care settings can use the presented steps to assess clinical workflow, thereby maximizing the match of their intervention with the clinical workflow. Conducting a prospective workflow study allows for evidence-based planning, identifying potential pitfalls, and increasing stakeholder buy-in and engagement. This tutorial should aid investigators in increasing the successful implementation of eHealth interventions.

(JMIR Mhealth Uhealth 2021;9(3):e18534) doi:10.2196/18534

KEYWORDS

workflow; implementation science; primary care; eHealth; stakeholder engagement

Introduction

Background

Within health care settings, one of the most common reasons eHealth apps fail to effectively increase the health outcomes they are designed to aid and improve quality of care is incompatibility between the app and clinical workflow [1-7]—the series of tasks conducted to complete clinical care in what order and by whom. Incompatibilities often stem from a wait-and-see method of implementation common to many eHealth interventions whereby an intervention is introduced into a clinical setting and the clinical workflow either adjusts to accommodate or does not [8,9]. Frequently, a lack of integration with clinical workflow results in clinical staff creating work-arounds or adaptations that interfere with the core components of the intervention [2-4,10]. For example, nurses reported checking boxes on forms that were not applicable to move the screens forward and others reported supplementing the electronic records with handwritten notes on paper [10]. Many eHealth apps would be more effective if a tailored implementation plan was created based on a systematic clinical workflow analysis performed before implementation [11].

Several methods for assessing clinical workflow are available [12-17]. Most notably, the Agency for Healthcare Research and Quality (AHRQ) has a digital toolkit of resources to guide investigators in studying clinical workflow [12]. In addition, Ozkaynak et al [13] have written an overview chapter of clinical workflow, including definitions of workflow, several workflow evaluation approaches, and examples of visualizing workflow [13]. Other studies describe select methods of studying clinical workflow, including electronic health record (EHR) audit logs, direct observation by trained external staff, clinic staff reporting, and sensor-based recordings [14-17]. Despite evidence that an intervention's compatibility with the local setting is pertinent to success [18], there is limited information clearly describing how to increase compatibility with clinical workflow before implementation within the eHealth literature.

According to the Consolidated Framework for Implementation Research (CFIR), a well-known implementation research framework, interventions are not inherently compatible with specific settings, and adaptation is required to maximize success [19,20]. Following this meta-theory, interventions can be adapted as long as the changes occur within the adaptable periphery—components that do not compromise the intervention's core components essential for the affect. One type of adaptation to improve user compatibility, a key construct from the diffusion of innovations theory incorporated within the CFIR inner setting [19,21,22] that can frequently be made without affecting the intervention's core components, is integration with the clinic's workflow. To ensure that adaptations enhance compatibility and remain within the adaptable periphery, it is suggested that the investigators and clinic staff collaborate to develop adaptations. In addition to engaging the clinical staff stakeholders, a key construct of CFIR and a critical component of integrating eHealth into primary care, this collaboration increases the likelihood of meeting two additional constructs within the CFIR process domain: identifying clinical champions and garnering the support of opinion leaders [19,22,23].

Objectives

This paper offers a simple-to-follow methodology for studying clinical workflow and planning for implementation with clinic staff. It provides tools adapted from previously published strategies to aid researchers in identifying clinic-specific adaptations that will increase the compatibility of eHealth interventions with clinical workflow and thus improve outcomes [24,25]. We present 4 steps for assessing clinical workflow and improving intervention compatibility with clinical settings, and a case study illustrating these steps. First, identify which components of the intervention are critical and when these components need to occur during a clinic visit. Second, choose from a variety of described methods to observe the existing clinical workflow relevant to the intervention. Third, confirm their findings using a second workflow assessment strategy. Finally, consult the clinic staff on the best ways to adapt the intervention compatibility with the confirmed clinical workflow. The purpose of this paper is to provide a tutorial for eHealth researchers on how to assess clinical workflow and use the knowledge gained to adapt interventions for maximal implementation success.

Assessing Clinical Workflow

Workflow Assessment Steps

Clinical workflow assessments for implementation planning can follow 4 steps (Table 1): (1) the identification of discrete workflow components, (2) workflow assessment, (3) triangulation, and (4) the stakeholder proposal of intervention implementation.

Table 1. Steps for conducting a workflow assessment.

Step	Purpose	Methods	Example tools
Identify discrete workflow components	Define what is necessary to make the intervention work	Select locations, interactions, and tasks	Review direct observation checklist (Multimedia Appendix 1) and other checklists to select clinical practices [26]
Workflow assessment	Create a model of the clinical workflow	Direct observation, interviews, sensor-based investigations, EHR[a] audit logs, and job task diaries	Review direct observation form (Multimedia Appendix 1)
Triangulation	Confirm rigor of clinical workflow model	Direct observation, interviews, sensor-based investigations, EHR audit logs, and job task diaries	Review semistructured template (Multimedia Appendix 2)
Stakeholder proposal	Plan intervention implementation based on stakeholder preferences	Interviews	Review semistructured template (Multimedia Appendix 2)

[a]EHR: electronic health record.

Step 1: Identification of Discrete Workflow Components

The first step in conducting a clinical workflow analysis for improving eHealth implementation planning is to identify the discrete components of the clinical workflow that need to be measured. Owing to the multitude of tasks occurring during clinical visits, it is imperative to clearly define what activities researchers should track to ensure that each is collected and documented consistently. Tracked activities may fall under 3 observable workflow categories: location, interactions, and tasks. Location refers to where and how long individuals are physically present in specific areas of a clinic. Interactions include face-to-face conversations and moments in which health records or patients transition from one provider to another. Tasks include a review of health records, measures collected (eg, weight, blood pressure), interventions administered (eg, giving prescription or vaccine), and administrative functions (eg, patient check-in or scheduling). For each task considered, the actors (patient, provider, or support staff) may vary.

To identify the discrete workflow components for the intervention, consider what is necessary to make the intervention work as intended in the setting. Imagine how the intervention will be integrated into the clinic and considered the following questions: At what time during a clinic visit is the most useful for the patient and provider to access the intervention? What specific tasks need to occur to have the intervention in the hands of those who need it at the right time during the clinical encounter? What types of clinical situations require alternative planning? When choosing the discrete workflow components of interest, it is important to consider the minimum necessary rule and evaluate tasks directly relevant to an intervention's implementation that do not compromise patient privacy or impede care delivery.

Step 2: Workflow Assessment

A second step in conducting a workflow analysis to improve eHealth intervention implementation is to assess the clinical workflow. Several strategies can be used to assess clinical workflow, including direct observation, clinical staff reporting, interviews with staff or patients, sensor-based investigations, EHR audit logs, and job task diaries [14,16,27-31]. We focused on direct observation by a member of the research staff because this method offers advantages over other possible choices, including an in-depth analysis of complex interactions and clinical work-arounds, limited equipment and technology needs, and the potential to increase clinical staff engagement [5,16,29].

Workflow assessment via direct observation can occur in 3 steps: (1) identify who you wish to observe, (2) complete informed consent with potential participants, and (3) directly observe the workflow using standardized data collection tools to systematically record observations. Before direct observation, engage with clinic leadership to ensure that the observed activities will not affect medical staff employment. Before the observation days, explain the informed consent to the identified staff so that they can thoughtfully consider participation. Observations should be scheduled for clinic days in which relevant patients have scheduled visits. Once patients express interest to clinical staff, patient consent can be obtained. Care must be taken to avoid disrupting the standard clinical flow to maintain high-quality research and clinical care. Although the number of observed patients can be influenced by practical considerations (eg, travel expenses and clinical receptiveness), to achieve a valid sample, patients should be observed until additional observations no longer present new information (ie, saturation is achieved) [32,33].

The use of a standardized observation form increases the rigor and reproducibility of workflow observations and can be incorporated into any workflow assessment method chosen [15]. The standardized observation form should include a simple method to record the location and primary actors for each of the discrete workflow components of interest. Multiple data collection forms and observers may be necessary because clinics have several actors attending to each patient simultaneously. Examples of standardized observation forms, including an adaptable Microsoft Access database, can be found in the AHRQ health information technology workflow assessment toolkit [34]. Including timestamps of observed activities in the standardized observation form enables researchers to conduct what is considered the gold standard measurement for clinical workflow assessment, a time and motion study [29]. Time and motion studies evaluate the activities and duration of each activity of the clinical workflow.

Step 3: Triangulation

A third step in conducting a clinical workflow analysis for improving eHealth intervention implementation is method triangulation: the verification of findings from a different viewpoint by answering similar questions with a different method [35]. Triangulation is a common strategy for enhancing the rigor of qualitative studies [32]. If triangulation reveals small deviations, the workflow should be updated. If triangulation reveals large deviations, additional data collection should be considered.

All the options mentioned for primary workflow assessment are possible methods for triangulation [16,27-31]. An additional option for triangulation is member checking of results, whereby the observed workflow is presented visually to the stakeholders via a sequential task analysis, line graphs, flow charts, and diagrams [13,17,36,37]. The AHRQ provides several examples of how to present workflow observations as flowcharts [38].

Step 4: Stakeholder Proposal of Intervention Implementation

It is useful for stakeholders or actual clinical users to evaluate how the intervention will be compatible with their clinical workflow. Stakeholder opinions on implementation can be obtained during the triangulation phase through interviews or planning meetings. The planning process should follow the community participatory principles of engagement and participation so that the clinic staff feel that their insights and expertise are critical to the project [39]. Examples of critical research staff attitudes include mutual respect and genuineness, transparent processes, and balancing power [39]. Critical components for the success of stakeholder engagement include genuine partnerships, strategic selection of clinicians at each site to be involved in the project, and accommodation of stakeholder needs (eg, patient care responsibilities) [40]. A framework for operationalizing the engagement of partners, including examples of practices from research projects, is available from the Patient-Centered Outcomes Research Institute [41]. Finally, research staff should use this opportunity to discover or gain understanding into further adaptation needed for special situations (eg, acute visits vs wellness visits).

Case Study

Overview: Protect Me 4

We developed a tablet-based eHealth app for primary care clinics called *Protect Me 4*. The app is intended for use by parents of adolescents at the beginning of their child's appointment to help parents learn about recommended vaccines and explore educational information addressing common hesitations. At the time of the appointment, the app also prompts providers of due vaccines and parents' selected hesitations to prepare the provider to lead more efficient and effective vaccine discussions.

Our pilot trial demonstrated that the intervention successfully increased initiation of our main target, human papillomavirus (HPV) vaccine [6]. The app's overall effectiveness was limited because only 8% (57/1062) of the eligible families used our app. On the basis of poststudy interviews with clinic staff, we hypothesized that the clinic workflow was not aligned with the implementation strategy. Thus, before our second study with an enhanced version of the app, we conducted a workflow analysis at 4 participating primary care clinics to understand how we could adapt our intervention to improve compatibility. All activities were approved by the University of Florida Institutional Review Board.

Step 1: Identification of Discrete Workflow Components

We identified 3 workflow-related conditions necessary for the success of the intervention. First, a parent needed 5-10 minutes to complete the application before the child saw the provider. Second, the provider needed to view the app results before (or during) the time spent with the patient. Third, the ordering and administration of vaccines needed to occur following parents' and providers' app use. As such, we needed to identify a time during a visit when the parent could complete the intervention, a method for getting the completed intervention to the provider for review, and ensure that both steps took place before the time when vaccines need to be ordered during a clinical encounter. Thus, we selected the following discrete workflow components of interest: (1) patient locations, interactions, and tasks before meeting with the provider; (2) provider interactions with the patient and patient records; (3) the timing of vaccine record review; and (4) the timing of vaccine administration. We restricted our workflow components to activities that occurred outside the clinic exam rooms in order to preserve patient privacy, encourage patient participation, and follow the minimum necessary rule for data collection.

Step 2: Workflow Assessment

To enhance the rigor of our direct observations, we created 2 standardized checklists by adapting the Arkansas Foundation for Medical Care Workflow Assessment Checklist found on the AHRQ website [34]. The first checklist (Multimedia Appendix 1) standardized recording observations of the patient from check-in to check-out and included fields about (1) patient locations, interactions, and tasks before meeting with the provider; (2) the timing of vaccine record review; and (3) the timing of vaccine delivery. The second checklist standardized the recording observations of the provider and included fields on provider interactions with patient records.

We created checklists in the MediDocs platform (EnMedical Systems), a software designed to capture medical information that includes a click to timestamp functionality. To ensure the usability and functionality of the Medidocs platform and enhance observer comfort with the procedures, observers created an example list of coffee shop procedures and then pilot tested the Medidocs platform at a local coffee shop by tracking employee and customer interactions. A coffee shop was selected because there were many individuals who followed the same procedures, and personal health information was not included.

We observed the clinical workflow at 4 primary care clinics that agreed to participate in the intervention implementation trial. The participating clinics were pediatric or family medicine specialties, served a range of patients with Medicaid insurance (14%-90% of patients), and were owned privately, by the

university, or by health systems. Before the observation days, a lead clinical stakeholder at each clinic invited all medical staff who interacted with patients aged 11-12 years to attend an information session. Information sessions were held during each clinic's preferred schedule (ie, during lunchtime or a staff meeting), food was provided, and research staff explained the study and informed consent forms. Clinic staff were given the opportunity to ask questions and return signed informed consent forms at any time before the observation days. The direct observation of clinical staff was restricted to those staff members who had returned signed consent forms.

To minimize clinic disruption and research staff travel expenses, we aimed to observe approximately 3 participants over 1-2 days at each clinic. Clinical staff selected days when at least one patient aged 11-12 years was scheduled. On each observation day, clinic staff invited all parents of patients aged 11-12 years to participate and referred interested parents to research staff who were present in the clinic. The research staff obtained consent through a Health Insurance Portability and Accountability Act waiver of documentation of consent and collected data without identifiers. Research staff conducting observations had a minimum of a bachelor's degree in a scientific or social science field, had existing relationships with the clinic, and were trained in research and practice facilitation. We chose not to interview patients about the acceptability of completing the eHealth intervention during a clinic visit because parents were very receptive to completing the eHealth intervention during our pilot, and we enhanced parents' acceptability of the app content with focus groups [6,42].

With 12 patients as a practical target, we observed 13 patient visits between August 2016 and March 2017. By the 13th visit, we were no longer observing new variations in the workflow components. Thus, saturation was reached, and further data collection was not needed [33]. However, we did continue to observe a variety of wait times but were able to conclude that sufficient time was typically available before the patient saw the provider.

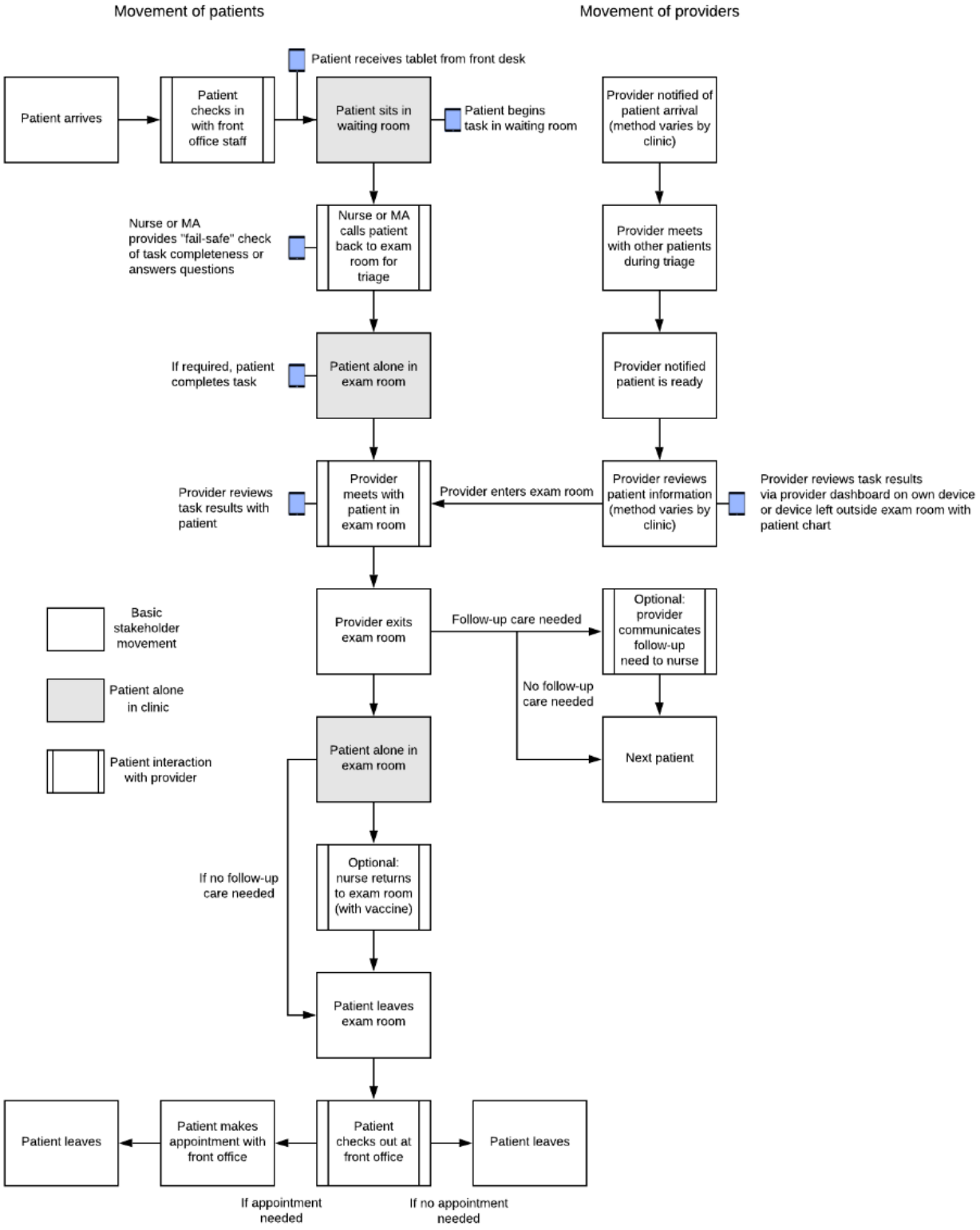
Step 3: Triangulation

For the triangulation of our direct observation, we interviewed clinical stakeholders (group interviews of a clinical support staff member and the lead physician at each clinic) because involving the clinic staff in the development process creates a sense of ownership and increases buy-in consistent with the CFIR [5,19,22]. We chose to visually present the workflow using a common flowcharting strategy [24]. The moderator was the same research staff who conducted direct observations. Each group followed a semistructured interview guide (Multimedia Appendix 2). To review the observed workflow, flowcharts of the acute and preventive care workflows were displayed on large poster boards. A moderator explained the flowchart, asked about any missing components, and solicited opinions on accuracy. Clinic staff were compensated US $20 for interview participation in approximately 20-minute sessions. Interviews were recorded and transcribed. The flowchart was updated for minor deviations. Evaluation activities used a postpositivist approach to observe and understand the main activities involved in clinic appointments.

After noting considerable similarities across clinics and visit types, we created a composite flowchart that represented the workflow at all 4 clinics (Figure 1). Our example flowchart shows that patients first arrived and were checked in with front office staff. Then, patients spent up to 59 minutes in the waiting room without staff contact (mean 12.6 min, SD 16.1 min in addition to time spent completing the study consent). In the 2 clinics, wait time was substantially less for acute than preventive visits. Next, a nurse or medical assistant triaged the patients' presenting issue and collected basic vital information. Patients were then left alone in the exam room for a range of 2-25 minutes (mean 10.9 min, SD 8.8 min).

While a patient completed check-in and triage, physicians attended to other patients. Once notified that a patient was ready and waiting in an exam room, physicians reviewed the patient's record for a few minutes immediately before entering the exam room. Depending on the clinic, patient records were reviewed via paper files placed immediately outside an exam room or electronically. Following the patient-physician interaction, physicians at each clinic communicated verbally with a nurse or medical assistant about additional care needs (eg, vaccine administration). The clinical staff administered ordered services. Finally, the patient returned to the front desk and scheduled another appointment or received the required instructions.

Figure 1. Example clinical workflow diagram of patient visits (those aged 11-12 years) with the proposed intervention tablet placement. MA: medical assistant.



Step 4: Stakeholder Proposal of Intervention Implementation

Clinicians suggested additional staff double-checking that the tablets were distributed. Participants mentioned that checking for tablet distribution during triage would be compatible with redundancy checks already taking place within a clinical workflow. “We always double-check ages and things anyway...so if I see that the patient is 11 or more, I can always grab [the tablet].” Another site noted that it would be helpful to use a team approach to enhance intervention dissemination. To accomplish this, front office staff could highlight the patients eligible for the intervention during their morning patient reviews. For example, one clinician stated:

> *They [front desk staff] identify all the 11- and 12-year-olds coming in for the following day, and they somehow highlight, red flag, that particular*

> *patient, and when we discuss patients coming in...it can just be mentioned. So, everyone, front desk staff, and clinicians, will know that a patient needs an iPad.*

At all 3 of these clinics, physicians preferred to review the intervention results at the same time they reviewed the patient records. Clinicians suggested physically placing the tablet with the encounter forms and patient records or accessing the intervention results on their computer when reviewing patient records. As one physician explained, "Usually I look through the patient chart before I walk into the room, so I would just do that [review intervention results] right after." In cases where patients were unable to complete the intervention in the waiting room, physicians felt that the parent could continue to complete the tablet while the physician interacted with the adolescent. In explaining this alternative, a physician noted:

> *...at age 11 or 12, I do more interaction with the child versus interaction with the parent. So, I can say, "Oh, by the way mom, can you go ahead and fill this out for me?"*

One clinic that already used a tablet system suggested a slightly modified implementation. The staff of this clinic proposed that the new intervention be distributed in a similar fashion to their current tablet system (ie, given by nurse to parent in the exam room). If the parent finished before the nurse completed triage activities and left the exam room, then the nurse would collect the tablet from the parent and place it outside the exam room for the physician to review. If a parent was still completing the tablet intervention during the completion of triage activities, the nurse would alert the physician that the parent still had the tablet. In this case, physicians reviewed the tablet results upon entering the room alongside the patient. One physician explained:

> *I'd bring it up and say, "Let me see how you feel about the HPV vaccine," and if they are ready to have it, "great," and if not, I would discuss if they had a hesitation.*

Clinicians identified 3 potential barriers to intervention implementation. First, clinicians reported that they do not typically review and administer vaccines during acute care visits. Although it was noted that parents would still be amenable to completing the intervention during most acute care visits, it was less clear how medical staff could alter their work processes to incorporate speaking about and, potentially, administering vaccines during acute visits. This issue was first mentioned in relation to the lack of an established process for checking vaccine records for patients presenting for an acute care visit, with staff reporting, "We don't usually check ahead for the sick visit. Just for the wells."

Second, physicians mentioned patient conditions that would preclude their administration of vaccines (eg, a high fever) or cause disruptions to how patients move through the clinic, thus interfering with when a tablet could be given to a patient to complete the intervention. One clinician explained that "...if we know that [a patient] is really, really contagious, we won't even do two minutes [in the waiting room], as soon as they come in we put them in an empty room." Clinicians were open to the idea of administering the intervention in these acute visit scenarios, for example, stating, "We could implement [the tablet] at some point, like once we've discussed the issue or whatever they've come in for," but noting that more work would need to be done in order to optimize the process for acute visits.

The final barrier mentioned was the physician time constraints. One physician mentioned that, despite the simplicity of the tablet, competing demands may cause him to forgo reviewing intervention results stating, "If I'm in a rush, or I'm behind, or if we have lots of add-ons, what have you...I mean, it's simple enough to just log-in and everything, but it's that extra step."

Discussion

Implications

We present a general strategy and an example for conducting a workflow assessment to enhance the implementation of an eHealth intervention in primary care. With this strategy, we were able to identify specific timing, staffing, and management processes that may enhance intervention implementation. Potential barriers related to the implementation of the intervention were identified and addressed by stakeholders. Incorporating clinical stakeholders into the design of the implementation plan enhanced stakeholder buy-in, a critical component of successful implementation of eHealth interventions [5]. By conducting the workflow analysis before the implementation of an eHealth intervention, researchers may avoid potential implementation barriers that limit the reach and effectiveness of the intervention.

Conducting the workflow assessment before intervention implementation allows for adaptations consistent with the adaptable periphery of interventions in the CFIR. By using a direct observation approach to assess clinical workflow, we were able to identify how and when our eHealth intervention should be incorporated. Clinicians' participation in verifying the workflow and proposing an implementation strategy allowed for additional options to be proposed before implementation. For example, clinicians added a double-check of the tablet distribution and provided contingency plans for when parents did not have enough time to finish the app before seeing the provider. Although the clinicians were unaware that our prior pilot had issues with front office staff forgetting to give the tablets to parents, incorporation of clinicians in the planning process pre-emptively addressed this shortcoming in our second implementation. Moreover, while we had anticipated a workflow tailored to each clinic based on prior research [37], conducting the workflow assessment allowed us to simplify our approach with a standardized implementation plan that could be applied to all clinics with minimal tailoring.

By engaging stakeholders with implementation planning, we were able to identify potential barriers, discourage clinic staff adaptation of the core intervention, and build clinician buy-in. Consistent with previous primary care workflow research [37], we found that the clinical workflow differed between acute and well visits, suggesting that it will be more challenging to use the app during acute visits. However, clinicians suggested work-arounds for acute visits that were inconsistent with the core of the app. For example, parents using the app after their

initial discussion with the provider allows clinicians to address immediate concerns but does not allow the app to aid parent-provider conversations. Learning of this potential work-around before implementation allowed research staff to explain the inconsistencies between the work-around and the core of the intervention and work with clinicians to develop alternative solutions (eg, limiting use of the app to nonurgent acute visits). Finally, and most importantly, through their involvement with the planning, clinicians gained a sense of ownership, which is expected to translate to a more successful implementation.

Limitations

There are 3 important limitations of the case study presented that could be improved in future workflow assessments. First, the workflow assessment focused on physicians and patients. Although physician and patient visits are the most frequent opportunity for administering adolescent vaccines, a workflow assessment that included nurses and other clinical staff could be useful if the intervention was implemented in nurse-only visits. Second, our triangulation and stakeholder engagement strategies focused on physicians and clinical support staff. In particular, as front office staff were identified as integral to the implementation plan, including front office staff in the workflow assessment and intervention implementation planning would likely improve planning and garner the needed support from these critical stakeholders. Third, our case study presented 4 clinics with 3 observations each and, in turn, the small number of observations increases the risk of undue influence of outliers if the visits are not representative of standard care practices. However, this size is reasonable for a small implementation study and is supported by thematic saturation [43].

Strengths

Our proposed workflow assessment strategy and supporting case study have 3 strengths relative to the design and execution. First, the use of 2 strategies to assess workflow enhances the rigor and validity of the results. Second, the use of separate observers for the patient and the physician was relatively novel and ensured that overlapping activities were observed. This strategy of one observer to each actor can be expanded to any clinical setting. Third, the benefit of involving the clinic staff in implementation planning with regard to stakeholder ownership and engagement cannot be understated. Participating clinicians are interested in seeing the intervention succeed and are aware that some work-arounds may alter the core of the intervention.

Conclusions

Conducting a prospective workflow assessment can enhance the acceptability and performance of eHealth interventions in clinical practice. We presented a 4-step strategy to increase the compatibility between eHealth interventions and clinical workflow, as compatibility is one of the main predictors of clinician use [7,44]. Multiple options are presented for each of the 4 steps: (1) the identification of discrete workflow components, (2) workflow assessment, (3) triangulation, and (4) the stakeholder proposal for intervention implementation. By following the included step-by-step guide, eHealth researchers can study clinical workflows to design eHealth implementation strategies that complement rather than compete with clinical care. More importantly, through engagement of clinical staff in implementation planning, research staff can increase the clinical staff buy-in to the intervention and are able to stress that adaptations, while possible, need to be implemented only after considering the effects on the core intervention.

Acknowledgments

The authors thank Jennifer Mauck (Master of Public Health) and Lori M Drum (Licensed Mental Health Counselor and Certified Clinical Research Coordinator) for their assistance with data collection. They also thank Shivani Desai (Master of Public Health) for manuscript formatting assistance. This research was funded by the National Institutes of Health, R21 CA202011 (principal investigator: SS).

Authors' Contributions

SS led all aspects of the project. ES, MM, MV, LT, and SS recruited the clinics. SS, NR, ES, LT, MV, and MM planned the study. ES and MV performed direct observations and conducted semistructured interviews. SS and JT analyzed and interpreted the data and drafted the manuscript. JT, NR, ES, and SS created the tables and figures. All authors have reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Direct observation checklist for patients and clinic staff. Boxes above and below the tables create timestamps upon clicking.
[PDF File (Adobe PDF File), 330 KB - mhealth_v9i3e18534_app1.pdf]

Multimedia Appendix 2
Semistructured interview questions and subquestions were used to evaluate workflow observations.
[PDF File (Adobe PDF File), 97 KB - mhealth_v9i3e18534_app2.pdf]

References

1. Trivedi MH, Kern JK, Marcee A, Grannemann B, Kleiber B, Bettinger T, et al. Development and implementation of computerized clinical guidelines: barriers and solutions. Methods Inf Med 2002;41(5):435-442. [Medline: 12501817]
2. Liberati EG, Ruggiero F, Galuppo L, Gorli M, González-Lorenzo M, Maraldi M, et al. What hinders the uptake of computerized decision support systems in hospitals? A qualitative study and framework for implementation. Implement Sci 2017 Sep 15;12(1):113 [FREE Full text] [doi: 10.1186/s13012-017-0644-2] [Medline: 28915822]
3. Djalali S, Ursprung N, Rosemann T, Senn O, Tandjung R. Undirected health IT implementation in ambulatory care favors paper-based workarounds and limits health data exchange. Int J Med Inform 2015 Nov;84(11):920-932. [doi: 10.1016/j.ijmedinf.2015.08.001] [Medline: 26296433]
4. Yeung MS, Lapinsky SE, Granton JT, Doran DM, Cafazzo JA. Examining nursing vital signs documentation workflow: barriers and opportunities in general internal medicine units. J Clin Nurs 2012 Apr;21(7-8):975-982. [doi: 10.1111/j.1365-2702.2011.03937.x] [Medline: 22243491]
5. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
6. Staras SAS, Vadaparampil ST, Livingston MD, Thompson LA, Sanders AH, Shenkman EA. Increasing human papillomavirus vaccine initiation among publicly insured Florida adolescents. J Adolesc Health 2015 May;56(5 Suppl):40-46 [FREE Full text] [doi: 10.1016/j.jadohealth.2014.11.024] [Medline: 25863554]
7. Sedlmayr B, Patapovas A, Kirchner M, Sonst A, Müller F, Pfistermeister B, et al. Comparative evaluation of different medication safety measures for the emergency department: physicians' usage and acceptance of training, poster, checklist and computerized decision support. BMC Med Inform Decis Mak 2013 Jul 29;13:79 [FREE Full text] [doi: 10.1186/1472-6947-13-79] [Medline: 23890121]
8. Yadav K, Chamberlain JM, Lewis VR, Abts N, Chawla S, Hernandez A, et al. Designing real-time decision support for trauma resuscitations. Acad Emerg Med 2015 Sep;22(9):1076-1084 [FREE Full text] [doi: 10.1111/acem.12747] [Medline: 26300010]
9. Cady RG, Finkelstein SM. Mixed-methods approach for measuring the impact of video telehealth on outpatient clinic triage nurse workflow. Comput Inform Nurs 2013 Sep;31(9):439-449 [FREE Full text] [doi: 10.1097/01.NCN.0000432126.99644.6c] [Medline: 24080753]
10. Cresswell KM, Mozaffar H, Lee L, Williams R, Sheikh A. Workarounds to hospital electronic prescribing systems: a qualitative study in English hospitals. BMJ Qual Saf 2017 Jul;26(7):542-551. [doi: 10.1136/bmjqs-2015-005149] [Medline: 27129493]
11. Vigoda MM, Gencorelli F, Lubarsky DA. Changing medical group behaviors: increasing the rate of documentation of quality assurance events using an anesthesia information system. Anesth Analg 2006 Aug;103(2):390-395. [doi: 10.1213/01.ane.0000221176.27215.20] [Medline: 16861422]
12. What is workflow? Agency for Healthcare Research and Quality. URL: https://digital.ahrq.gov/health-it-tools-and-resources/evaluation-resources/workflow-assessment-health-it-toolkit/workflow [accessed 2021-01-19]
13. Ozkaynak M, Unertl K, Johnson S, Brixey J, Haque S. Clinical workflow analysis, process redesign, and quality improvement. In: Clinical Informatics Study Guide. Cham, Switzerland: Springer International Publishing; 2016:135-161.
14. Rule A, Chiang MF, Hribar MR. Using electronic health record audit logs to study clinical activity: a systematic review of aims, measures, and methods. J Am Med Inform Assoc 2020 Mar 01;27(3):480-490 [FREE Full text] [doi: 10.1093/jamia/ocz196] [Medline: 31750912]
15. Fontanesi J, De Guire M, Chiang J, Holcomb K, Sawyer M. Applying workflow analysis tools to assess immunization delivery in outpatient primary care settings. Jt Comm J Qual Improv 2000 Nov;26(11):654-660. [doi: 10.1016/s1070-3241(00)26056-2] [Medline: 11098428]
16. Ampt A, Westbrook J, Creswick N, Mallock N. A comparison of self-reported and observational work sampling techniques for measuring time in nursing tasks. J Health Serv Res Policy 2007 Jan;12(1):18-24. [doi: 10.1258/135581907779497576] [Medline: 17244393]
17. Vankipuram M, Kahol K, Cohen T, Patel VL. Toward automated workflow analysis and visualization in clinical environments. J Biomed Inform 2011 Jun;44(3):432-440 [FREE Full text] [doi: 10.1016/j.jbi.2010.05.015] [Medline: 20685315]
18. Pagel C, Bates DW, Goldmann D. How to avoid common pitfalls of health IT implementation. NEJM Catal Innov Care Deliv. 2020. URL: https://catalyst.nejm.org/doi/full/10.1056/CAT.20.0048 [accessed 2021-03-02]
19. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
20. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629 [FREE Full text] [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]
21. Rogers EM. Diffusion of innovations, 4th Edition. New York, NY: Simon and Schuster; 2010.

22. CFIR Research Team. Constructs. Consolidated Framework for Implementation Research (CFIR). 2021. URL: https://cfirguide.org/constructs/ [accessed 2021-01-15]
23. Bally ELS, Cesuroglu T. Toward integration of mHealth in primary care in the Netherlands: a qualitative analysis of stakeholder perspectives. Front Public Health 2019;7:407 [FREE Full text] [doi: 10.3389/fpubh.2019.00407] [Medline: 32010660]
24. Unertl KM, Weinger MB, Johnson KB, Lorenzi NM. Describing and modeling workflow and information flow in chronic disease care. J Am Med Inform Assoc 2009;16(6):826-836 [FREE Full text] [doi: 10.1197/jamia.M3000] [Medline: 19717802]
25. Workflow assessment for health IT toolkit. Agency for Healthcare Research and Quality. URL: https://digital.ahrq.gov/health-it-tools-and-resources/evaluation-resources/workflow-assessment-health-it-toolkit [accessed 2021-01-19]
26. Checklist. Agency for Healthcare Research and Quality. URL: https://healthit.ahrq.gov/health-it-tools-and-resources/evaluation-resources/workflow-assessment-health-it-toolkit/all-workflow-tools/checklist [accessed 2021-01-19]
27. Carayon P, Karsh BT, Cartmill RS. Job task diary. Prepared by the Center for Quality and Productivity Improvement, University of Wisconsin - Madison, under Contract No. HHSA 290-2008-10036C. Rockville, MD: Agency for Healthcare Research and Quality; 2010. URL: https://digital.ahrq.gov/sites/default/files/docs/workflowtoolkit/JobTaskDiary.pdf [accessed 2021-03-02]
28. Wu DTY, Smart N, Ciemins EL, Lanham HJ, Lindberg C, Zheng K. Using EHR audit trail logs to analyze clinical workflow: a case study from community-based ambulatory clinics. AMIA Annu Symp Proc 2017;2017:1820-1827 [FREE Full text] [Medline: 29854253]
29. Bratt JH, Foreit J, Chen PL, West C, Janowitz B, de Vargas T. A comparison of four approaches for measuring clinician time use. Health Policy Plan 1999 Dec;14(4):374-381. [doi: 10.1093/heapol/14.4.374] [Medline: 10787653]
30. Lopetegui M, Yen P, Lai A, Jeffries J, Embi P, Payne P. Time motion studies in healthcare: what are we talking about? J Biomed Inform 2014 Jun;49:292-299 [FREE Full text] [doi: 10.1016/j.jbi.2014.02.017] [Medline: 24607863]
31. Kannampallil T, Li Z, Zhang M, Cohen T, Robinson DJ, Franklin A, et al. Making sense: sensor-based investigation of clinician activities in complex critical care environments. J Biomed Inform 2011 Jun;44(3):441-454 [FREE Full text] [doi: 10.1016/j.jbi.2011.02.007] [Medline: 21345380]
32. Palinkas LA. Qualitative and mixed methods in mental health services and implementation research. J Clin Child Adolesc Psychol 2014;43(6):851-861 [FREE Full text] [doi: 10.1080/15374416.2014.910791] [Medline: 25350675]
33. Padgett DK. Qualitative methods in social work research, 2nd ed. Los Angeles, CA: Sage Publications Inc; 2008:1-304.
34. Work flow assessment checklist. Arkansas Foundation for Medical Care. 2005. URL: https://healthit.ahrq.gov/sites/default/files/docs/workflowtoolkit/Workflow_Assessment_Checklist.pdf [accessed 2021-01-19]
35. Flick U, von Kardoff E, Steinke I. A companion to qualitative research. London: SAGE Publications Ltd; 2004:1-448.
36. Creswell JW, Miller DL. Determining validity in qualitative inquiry. Theory Into Practice 2000 Aug;39(3):124-130. [doi: 10.1207/s15430421tip3903_2]
37. Holman GT, Beasley JW, Karsh B, Stone JA, Smith PD, Wetterneck TB. The myth of standardized workflow in primary care. J Am Med Inform Assoc 2016 Jan;23(1):29-37 [FREE Full text] [doi: 10.1093/jamia/ocv107] [Medline: 26335987]
38. Flowchart. Agency for Healthcare Research and Quality. URL: https://digital.ahrq.gov/health-it-tools-and-resources/evaluation-resources/workflow-assessment-health-it-toolkit/all-workflow-tools/flowchart [accessed 2021-01-19]
39. Rhodes SD, Tanner AE, Mann-Jackson L, Alonzo J, Simán FM, Song EY, et al. Promoting community and population health in public health and medicine: a stepwise guide to initiating and conducting community-engaged research. J Health Dispar Res Pract 2018;11(3):16-31 [FREE Full text] [Medline: 31428533]
40. Forsythe LP, Ellis LE, Edmundson L, Sabharwal R, Rein A, Konopka K, et al. Patient and stakeholder engagement in the PCORI pilot projects: description and lessons learned. J Gen Intern Med 2016 Jan;31(1):13-21 [FREE Full text] [doi: 10.1007/s11606-015-3450-z] [Medline: 26160480]
41. Sheridan S, Schrandt S, Forsythe L, Hilliard TS, Paez KA, Advisory Panel on Patient Engagement (2013 inaugural panel). The PCORI engagement rubric: promising practices for partnering in research. Ann Fam Med 2017 Mar;15(2):165-170 [FREE Full text] [doi: 10.1370/afm.2042] [Medline: 28289118]
42. Theis RP, Wells BA, Staras SAS. "I can be the Judge of What's Serious": a qualitative pilot study of parents' responses to messaging about side effects of the HPV vaccine. Matern Child Health J 2020 Apr;24(4):456-461. [doi: 10.1007/s10995-019-02856-8] [Medline: 31893386]
43. Guest G, Bunce A, Johnson L. How many interviews are enough?: An experiment with data saturation and variability. Field Methods 2006 Feb 01;18(1):59-82. [doi: 10.1177/1525822X05279903]
44. Jun S, Plint AC, Campbell SM, Curtis S, Sabir K, Newton AS. Point-of-care cognitive support technology in emergency departments: a scoping review of technology acceptance by clinicians. Acad Emerg Med 2018 May;25(5):494-507 [FREE Full text] [doi: 10.1111/acem.13325] [Medline: 28960689]

Abbreviations

AHRQ: Agency for Healthcare Research and Quality

CFIR: Consolidated Framework for Implementation Research
EHR: electronic health record
HPV: human papillomavirus

Edited by L Buis; submitted 20.04.20; peer-reviewed by D Pförringer, M Ozkaynak; comments to author 29.06.20; revised version received 16.09.20; accepted 22.02.21; published 31.03.21.

Please cite as:
Staras S, Tauscher JS, Rich N, Samarah E, Thompson LA, Vinson MM, Muszynski MJ, Shenkman EA
Using a Clinical Workflow Analysis to Enhance eHealth Implementation Planning: Tutorial and Case Study
JMIR Mhealth Uhealth 2021;9(3):e18534
URL: https://mhealth.jmir.org/2021/3/e18534
doi:10.2196/18534
PMID:33626016

Original Paper

Interplay of Support, Comparison, and Surveillance in Social Media Weight Management Interventions: Qualitative Study

Leanne Chang[1], PhD; Kaushik Chattopadhyay[2], EPH; Jialin Li[3], MD; Miao Xu[3], MD; Li Li[3], MD

[1]Department of Communication Studies, School of Communication, Hong Kong Baptist University, Hong Kong, Hong Kong

[2]Division of Epidemiology and Public Health, School of Medicine, University of Nottingham, Nottingham, United Kingdom

[3]Department of Endocrinology and Metabolism, Ningbo First Hospital, Ningbo, China

Corresponding Author:
Li Li, MD
Department of Endocrinology and Metabolism
Ningbo First Hospital
59 Liuting Street
Ningbo, Zhejiang 315010
China
Phone: 86 574 8708 5588
Email: lilyningbo@163.com

Abstract

Background: There has been a significant increase in the trend of using social media as a platform to deliver weight management interventions. This illustrates a need to develop a holistic understanding of doctor-patient communication and peer-to-peer communication in social media interventions and to determine their influences on weight management for people with overweight or obesity. Such studies will highlight how social media can be more effectively integrated into weight management programs to enhance individuals' short-term and long-term weight management behaviors.

Objective: The aim of this study was to examine patients' experiences with doctor-patient communication and peer interactions in a social media–based (WeChat) weight management program, and to describe the interplay of three social influence factors—social support, social comparison, and surveillance—in their weight control practices. The program, designed and implemented by the research team located in a tertiary referral hospital in a southeastern province in China, included both diet and physical activity components that targeted people with overweight or obesity.

Methods: We conducted in-depth interviews with 32 program participants of different ages (mean 35.6, SD 7.7 years), gender (18 women), duration of program membership (mean 1.4 years), and weight loss outcomes (54% weight loss to 9% weight gain). All interview data were audio-recorded, transcribed, and translated using the translation-backtranslation technique. Nvivo software was used to facilitate the coding process.

Results: Results of thematic analysis indicated the distinct functions of professionally led support and peer support. Professional support was presented in the form of knowledge infusion, efficacy enhancement, and provision of timely feedback. Peer support fostered empathy and sense of belonging, and had a mutually reinforcing relationship with peer comparison and peer-based surveillance. Peer comparison enhanced motivation and positive competition. However, it also reinforced negative group norms, and resulted in downturns in reference standards and collective inactivity. Social media surveillance prompted participants' reactions to the gaze from medical professionals and peers that could be encouraging or inhibiting. Surveillance enhanced vigilance with weight control norms; however, its influence weakened when participants chose to fake weight data and turn off notifications. Findings from this study illustrated the interrelated and fluctuating influences of support, comparison, and surveillance.

Conclusions: The interactive traits of social media eased the practices of social support and social comparison, and created new forms of surveillance. This study contributes to an in-depth understanding of social media influences on individuals' weight control behaviors. Practical implications of the study concern improved strategies for maintaining the positive dynamics of social media interactions and preventing negative resistance to surveillance technology.

Trial Registration: Chinese Clinical Trial Registry ChiCTR1900025861; http://www.chictr.org.cn/showprojen.aspx?proj=42497

(JMIR Mhealth Uhealth 2021;9(3):e19239) doi:10.2196/19239

KEYWORDS

obesity; social comparison; social media; social support; surveillance; weight control

Introduction

Background

In China, an estimated 32.3% of the adult population had overweight or obesity in 2016 [1]. This number has doubled since 1992, and China is now alongside the United States as the two nations with the largest percentage of citizens affected by overweight and obesity [2]. Standard obesity prevention and treatment focuses on improving individuals' self-regulation of dietary regimens and physical activity [3,4]. In recent years, there has been an increasing trend of integrating social media into weight management interventions to improve participant engagement and retention [5,6]. These social media platforms include blogs, discussion forums, social networking sites, and other online communities using Web 2.0 technologies permissive for user-generated content and information exchange [7,8].

Social media has demonstrated the potential to facilitate weight management through improving provider-patient communication and peer-to-peer communication in cyberspace [8,9]. A review of the literature indicated that integrating social media into weight management interventions may increase participants' direct interactions with health professionals and deepen provider-patient relationships [9]. Social media may enhance connections among users with common interests in weight control and amount to a new phenomenon of "peer-to-peer health care" [8]. A scoping review of social media in dietetic practices found that being able to ask for help from health professionals and perceived support from peers were important facilitators to participant engagement and dietary behaviors [6]. However, two systematic reviews of randomized controlled trials reached the same conclusion of no direct impact of social media use on weight outcomes [10,11]. One possible explanation is that social media may promote initial changes but cannot sustain them in the long run [11]. Hence, finer details are needed to understand the complex long-term influences of social media interactions [10].

Noting the complexities of social media dynamics, this study adopted a holistic approach to examining the influences of social media interactions on weight management. Specifically, this study examined three types of social influences—social support, social comparison, and surveillance—that may arise from two types of social media interactions: doctor-patient communication and peer interactions. By examining participants of a social media–based weight management program that has lasted for more than 2 years, this study contributes to addressing many of the challenges with interventions over short time frames [11]. The focus on a Chinese social media app used among a Chinese population where obesity is on the rise also adds new knowledge to the current body of literature on social media weight management interventions.

Social Support

Social support refers to resources available from one's social network that are intended to be helpful [12]. The support resources could be informational (eg, advice and knowledge), instrumental (eg, material and financial aid), emotional (eg, encouragement and empathy), and appraisal (eg, affirmation and evaluative feedback) [13]. Conceptual studies suggest that support from health professionals and peers differs in their functions and ways of promoting behavioral change [14,15]. Professionally led support may facilitate health behaviors through skills training, monitoring, and feedback that appear to be directive, prescriptive, and guided by rules [14]. Peer support may promote self-regulation through social norms, identity, and companionship, and appear to be noncompulsory, mutual, and decentralized [14].

Social support is not a new concept in health interventions. However, limited studies have examined both professional support and peer support in social media weight management interventions. One field experiment compared 425 users' submission of dietary diaries and found that professional support had positive effects on diary submission, whereas peer support curbed user participation [16]. The negative effect of peer support was attributed to the perceived threats from fellow users who reported better dietary practices [16]. Another clinical trial with 301 women controlled for the inclusion of professional support and peer support in a web-based weight management intervention. No group difference in weight change was found, although the professional support group showed longer retention than the peer support group [17]. A focus group study including 35 adults with overweight or obesity revealed that peer support provided companionship, while professional support provided tailored instructions. However, their impacts on weight control were unexplored [18]. These initial findings guide further investigation of the coexistence of the two types of social media support and their interplay in weight management.

Social Comparison

Social comparison involves one's comparison to similar others for maintaining a stable view of oneself [19]. The concept of online social comparison has been widely studied in the literature on social media and body image, highlighting its negative effects on weight concern and disordered eating [20]. Recently, more research has looked into the positive impacts of social comparison on weight loss [9]. An experiment on a Facebook exercise app indicated that when users were allowed to track each other's physical activity records and progress, they exercised more than those in the nonsocial conditions. The benefits of positive competition were emphasized in user feedback [21]. In a large-scale online weight management program, the design of within-team and among-team competitions improved participants' overall level of physical activity, although the specific mechanisms by which social comparison exerted these influences were unspecified [22]. A metareview of 26 studies of physical activity apps found that modeling, information sharing, and social networking may be

the underlying factors prompting online social comparison [23]. One suggestion derived from the metareview was to investigate different dimensions of social comparison behaviors and their relationships with other social factors [23]. An interview study of user perspectives on a fitness app revealed that social comparison motivated some users to self-improve and be physically active; however, it also demotivated others when they were unable to catch up with high performers [24]. These findings pave ways for further exploration of how social comparison intervenes in online social networking and influences weight management in possible positive and negative ways.

Surveillance

Surveillance refers to the collection and monitoring of personal details to manage a population [25]. Foucauldian-based interpretation often uses the panopticon metaphor to describe surveillance as a form of control conducted through self-awareness and self-monitoring [26,27]. The panopticon is a theoretical prison design that places the guard tower at the center and the prisoners at the surrounding cells; the prisoners' awareness of the gaze from the invisible guards and their internalization of the norms explain their self-regulated behaviors [28]. On social media, users are both observers and observees who can track each other's personal information and be surveilled [29]. The unilateral communication taking place on social media fosters new practices of participatory surveillance that involve "many watching many others" among users, which differs from the conventional form of surveillance involving a few people of authority watching many observees [26,30].

Limited studies have examined participatory surveillance in weight-related social media interventions. One case study described that using social media to report daily health routines such as exercise, diet, and blood glucose level encouraged users with chronic illness to persist in self-health management [31]. The presence of an audience and feedback via social media enhanced self-surveillance, and one's continual practices and reporting of health behaviors [31]. Two focus group studies of school children's use of fitness apps indicated that displaying a positive image under the peer gaze was a major motivator for fitness pursuits and constant self-monitoring [27,32]. Adolescent users reported pressure to conform to fitness norms and perform self-tracking with the awareness of being watched [27,32]. Critical scholars have raised concerns about the negative outcomes of surveillance, such as invasion of privacy and resistance [26,30]. However, surveillance may also evoke agency and better self-monitoring of health [29]. More empirical evidence is needed on surveillance in social media interactions.

Based on this background, in this study, we sought to explore the cybersocial influences of social media interactions on weight management behaviors among people with overweight or obesity. The research question was as follows: How do social media support, comparison, and surveillance affect weight management among people with overweight or obesity?

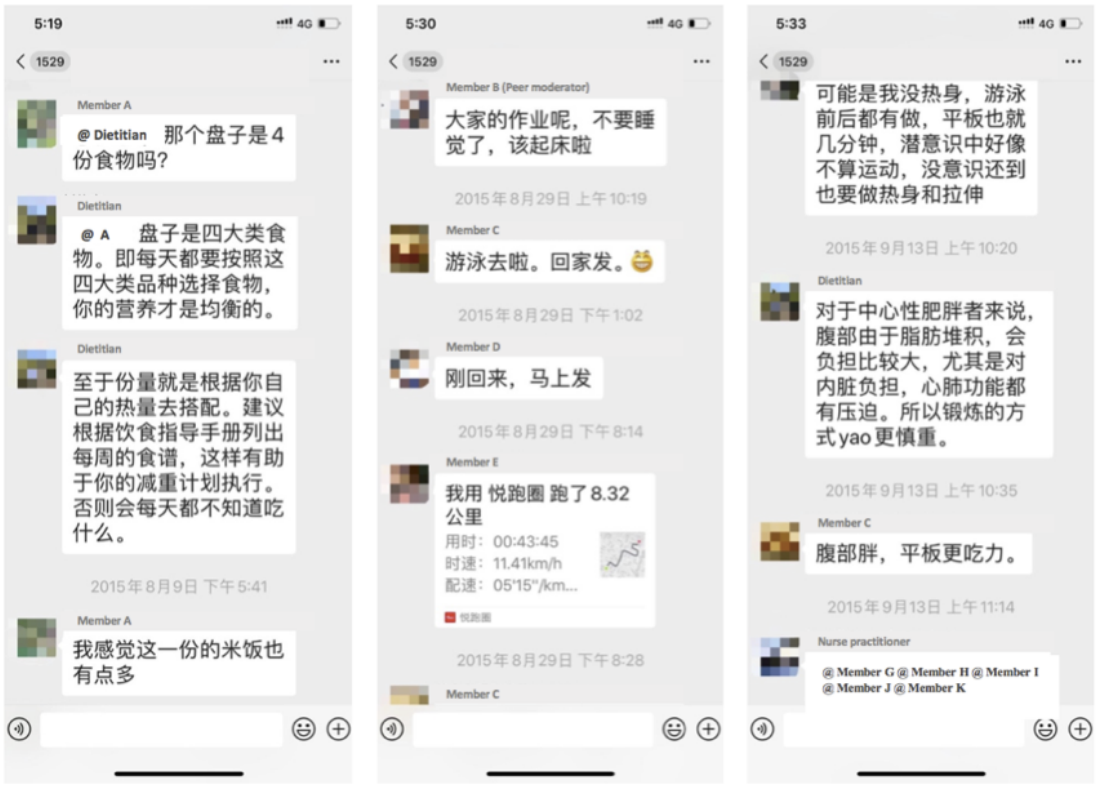
Methods

Research Context

The research context involved a free-of-charge WeChat-based weight management program implemented by a tertiary referral hospital in an urban city with a population of 8 million inhabitants in southeast China. WeChat is a multipurpose messaging, mobile payment, and social media app that functions similarly to the combination of Facebook and Apple Pay. By 2016, WeChat had reported a penetration rate of 93% in first-tier cities and an estimated penetration rate of 80% across the population in China [33,34]. Its prevalence made it an easy-to-use intervention delivery tool in the studied context.

Starting from July 2015, the program, which features scientific weight loss for people with overweight or obesity, was promoted through local television programs, newspapers, hospital websites, WeChat groups, and doctor recommendations during outpatient appointments. Interested individuals registered with the hospital and attended a weight management workshop to learn the basics of healthy eating (eg, nutrients and calories), exercise (eg, frequency, intensity, and duration), and program logistics in detail. Participants then enrolled in a WeChat private group with 10 to 12 members based on their joining time. An additional WeChat public group was made accessible to all participants. All participants used their extant WeChat account with their preferred display names to interact in the groups. Within each private group, a registered dietitian sent messages at agreed-upon times that asked participants to post their weight, dietary intake, and physical activity on a daily basis, and provided personalized feedback during the first month. Subsequently, participants were asked to report their weight on a weekly basis during the second and third months, and on a monthly basis from the fourth month onward. The three-phase design emphasized different goals (eg, from heightening awareness to weight loss maintenance) in different phases [35,36]. Two volunteer peer moderators assisted in sending reminders and collecting weight data from fellow members. A medical team of three registered dietitians, four endocrinologists, three orthopedists, and one nurse practitioner was introduced to the participants during the workshop and enrolled in all WeChat groups with their real names. All private and public groups remained open once established, wherein participants and medical professionals could freely interact with each other by posting content, commenting, and tagging a specific person to reply (Figure 1). At the time of the study, 28 private WeChat groups were created, with the oldest group being active for 2 years and the most recent for about 1 week.

Figure 1. Screenshots of interactions in a private WeChat group.



Member A: @ Dietitian Does the plate hold four servings of food?

Dietitian: @ Member A The [Healthy Eating] plate displays four basic food groups. You should choose food from each of the four food groups everyday so that your nutrition is balanced.

Dietitian: As for the portions, they depend on your caloric needs. I suggest that you list each week's meal plans based on the dietary guidebook, this will help you carry out your weight loss plan. Otherwise you won't know what to eat every day.

5:41 pm, 9 August 2015

Member A: I also feel the amount of rice [on the plate] is a bit too much.

Member B (Peer moderator): Where is everyone's homework? Stop sleeping, time to get up.

10:19 am, 29 August 2015

Member C: Went swimming. Will send it when I get home.

1:02 pm, 29 August 2015

Member D: Just came back, send it right away.

8:14 pm, 29 August 2015

Member E: I ran 8.32 km using RUN

Duration: 00:43:45
Average Speed: 11.41km/h
Average Pace: 05'15''/km

Member F: Probably because I didn't warm up. I did [plank] before and after swimming, just for a few minutes. In my subconscious mind, it wasn't counted as exercise. I wasn't aware that I had to warm up and stretch.

10:20 am, 13 September 2015

Dietitian: For people with central obesity, the accumulation of abdomen fat generates more burden, particularly to your internal organs, and cardiovascular function. So, the ways you exercise need to be more prudent.

10:35 am, 13 September 2015

Member C: Abdominal obesity makes it more difficult to do plank.

11:14 am, 13 September 2015

Nurse Practitioner: @ Member G @ Member H @ Member I @ Member J @ Member K

Procedures

All study procedures obtained ethical approval from the Research Ethics Panel at the University of Nottingham Ningbo China and the Research Ethics Committee at Ningbo First Hospital (Approval ID: 2017-R049). The interview participants were purposively selected with variations in gender, age, duration of membership, and amount of weight loss to allow for exploration of diverse experiences [37]. Recruitment was conducted through phone calls and face-to-face invitation during hospital visits in May and July 2017. All interviews were conducted in Mandarin by the first author between June and July 2017, and were scheduled at times convenient for the participants in a consultation room in the hospital. Participants were informed about the recruitment criteria, purpose of the study, interview process, nature of the questions to be asked, and their right to participate or withdraw on a voluntary basis. Each participant provided written informed consent prior to the

interview and received one hand towel set and a free body composition analysis at the end of the interview as a token of appreciation for their time. All participants completed the interview with the knowledge that they could stop at any time if they preferred to do so.

The interviews were semistructured, beginning with a short demographic survey and general questions such as "What were the reasons that motivated you to join the weight management program?" and "What health changes have you observed since joining the program?" Participants were then asked to share their weight control practices, perceived benefits and barriers to adhering to health suggestions, observations of peer interactions and doctor-patient communication in WeChat groups, and experience with support exchange among group members and support sharing from the medical team. Each interview lasted about 1 hour (mean 57.3 minutes, SD 7.0, range 45-75). After 32 interviews, the research team concluded that we had reached the point of data saturation after observing replications in the data and identifying no new ideas [38]. All interviews were transcribed verbatim by the first author and three student assistants in Mandarin, and then translated by four professional translators into English using the translation-backtranslation technique.

Participants

Thirty-two participants (14 men and 18 women) from 18 different WeChat groups completed the interviews. Their ages ranged from 21 to 53 years (mean 35.6, SD 7.7 years). Six participants served as peer volunteers and the rest were group members. The average duration of program participation was 1.4 years (mean 510.6 days, SD 203.4). The absolute weight change ranged from losing 55 kilograms to gaining 7 kilograms (mean 11.2 kilogram loss, SD 13.1). The percent weight change ranged from a weight loss of 54% to a weight gain of 9% (mean 12% loss, SD 13.3). Participants' demographics are shown in Multimedia Appendix 1.

Data Analysis

A thematic approach was used to identify themes [39]. The first author used Nvivo software to perform line-by-line analysis and generate initial codes. Through sorting, resorting, and iterative discussions, the first and second authors grouped interpretive categories into potential themes that addressed the research question. Finally, all authors examined the themes and agreed on their internal consistency and inclusion of explanatory accounts. Expert check, member check, and triangulation techniques were used to verify the validity of the results [40,41]. One dietitian and two participants reviewed the themes and agreed on the interpretations. The first author triangulated the results with 9 months of observations in the public WeChat group between August 2017 and April 2018 to validate the comprehensiveness of the findings.

Results

Overall Themes

Three themes were extracted to summarize the influences of social media interactions on participants' weight management: professionally led support for capacity building, cooperation and resistance to surveillance, and mutual reinforcement of peer influences.

Professionally Led Support for Capacity Building

Professionally led support was reflected in knowledge infusion, efficacy enhancement, and the availability of timely feedback. Table 1 presents the supporting quotes for this theme. Participants indicated that information from medical professionals corrected their misunderstanding about diet, exercise, metabolism, and ways to lose weight (quote 1). They shared common experiences of being exposed to unregulated drug advertisements, specious information about diet and exercise, and weight loss scams that circulated on different media or through personal networks and neighborhood pharmacies (quote 2). Without knowing the effects and side effects, they often tried these advertised drugs and weight loss packages, and learned their lessons only after feeling ill or paying money for products that did not work. Having a platform to communicate with medical professionals provided a reliable means to filter out false information and transform them "from amateur to academic."

Information and feedback from medical professionals helped participants set realistic expectations, assure certainty, and enhance their efficacy in coping with weight fluctuations (quotes 3-4). This support built participant confidence such as "as long as I keep myself on the right track, I will eventually lose weight." All participants shared that they had tried extreme weight loss methods such as consuming unregistered diet pills or fasting and observed no apparent effect. With support from medical professionals, they "became more rational" and could "stay calm even if I regain some weight."

The availability of timely feedback reflected another aspect of receiving professional support. Communicating via WeChat allowed participants to seek additional advice and clarification that was overlooked during hospital visits. Some minor problems could be solved quickly through real-time processing or asynchronous communication (quote 5). Participants were able to receive feedback within a few hours rather than waiting until the next doctor's appointment. Additionally, they could collectively learn from medical professionals or peer responses to avoid repetitive questioning (quote 6). The approachability of medical professionals also had a spillover effect of emotional support. WeChat communication shortened the relational distance between participants and doctors, thereby making the doctor-patient relationship "more like a friendship." When the doctors became friends rather than faceless figures, the strengthened bonds offered more emotional support and incentives for adherence (quotes 7-8).

However, professionally led support also had limitations, which were reflected in participants' use of knowledge gain as an excuse for procrastination, and the gap between knowing and doing. Once participants believed that they knew how to lose weight, the illusions may foster their decision to "take it easy" and "postpone the actual implementation of the process." Factors such as structural constraints (eg, family dietary habits and the business drinking culture), low perceived vulnerability, and hesitancy to curb old habits also discounted the impact of professionally led support (quotes 9-11).

Table 1. Supporting quotes for theme 1: Professionally led support for capacity building.

Subtheme	Quotes
Knowledge infusion	Quote 1: "She [dietitian] didn't give all the suggestions at once. Rather, she did this through day-to-day conversations, like making progress steadily in a quiet way. Some tips on a day and several on another, and in the end, we would bear in mind what is right and what is not."
	Quote 2: "There are advertisements about diet pills on the internet boasting about their effectiveness, but they don't mention the side effects. They claimed to be imported from Germany but, in fact, they were produced by small local food plants. There are many similar medicines on the internet. We can distinguish them by asking the doctors."
Efficacy enhancement	Quote 3: "The doctors would tell you that this was a normal bottleneck, especially after the early period of rapid weight loss. There'll be good results afterward. As everything is explained from a professional standpoint, it's easier to believe."
	Quote 4: "What I learned is a scientific and realistic approach to losing weight. Previously, we only knew that we should eat less, like dumping staple food. It worked, but things would go back quickly, and the effect just didn't last. They helped us develop new eating habits, and the biggest change is that I now know what food does me good and what does me harm."
Timely feedback	Quote 5: "If I forget to ask, I can ask them on WeChat and receive the answer soon. Otherwise, I have to wait until I meet the doctor again."
	Quote 6: "You can see how team members handled their situations…and the doctors did not need to repeat the same content to every group member."
Emotional support	Quote 7: "Our membership in the group is not like the doctor-patient relationship in a hospital. It is like becoming a friend to the medical team. Normally the doctor would wear a white gown, sitting there to examine you, but joining this group is different. The feeling is different."
	Quote 8: "Because you trusted them and could sense their sincerity, you would be more willing to follow their suggestions."
Inhibiting factors	Quote 9: "In China, people connect with each other often by having dinner or lunch together. It's the eating culture here. It's like people gather not to eat but to meet each other. I don't want others to realize that I'm thinking negative about the food. It's a bit tiring to keep doing so and it's easy to go back to old habits."
	Quote 10: "I have high cholesterol, blood pressure, and blood sugar, but I don't feel any physical pain. This leads to the fact that I fail to value the importance of losing weight. I have the knowledge but I'm in my twenties and you ask me to eat bland food. It's quite hard to do so."
	Quote 11: "I think if I completely refuse to eat what I like, that's a terrible life. Now I'm sort of thinking that a healthy mentality is much more important than a healthy lifestyle."

Cooperation and Resistance to Surveillance

WeChat created a surveillant environment that underpinned the perceived need to practice healthy eating and exercise behaviors. Table 2 presents the supporting quotes for this theme. The presence of medical professionals in the groups and an anticipation of their evaluations constantly reminded participants about being in the process of weight control and the need to self-regulate their behaviors (quotes 12-13).

However, since surveillance was performed in cyberspace, it lacked a mechanism to verify untruthful data. Five participants explicitly described the experience of reporting fake body weight and calorie intake to save face and avoid blame (quotes 14-16). When weight loss stalled, participants were lured into faking weight data with the hope that they could catch up to the falsely reported number at a later date. Once they failed to reach the desired figures, frustration and embarrassment could drive them to hide their weight loss outcomes, either by staying silent in the groups or repeating false information to make their weight data look acceptable. In most of these cases, insignificant weight loss and disengagement became a vicious circle of ineffective weight management.

Resistance to surveillance was also present in participants' reactions to frequent WeChat notifications. Participants wanted to attend to group messages and gain information. A vibrant online community encouraged their engagement. However, message posting, replying, and tagging could be initiated by anyone at any time (Figure 1). This resulted in a large number of push notifications that were perceived to be irrelevant and disturbed participants' daily activities or rest (quotes 17-18). Many participants reported the experience of neglecting all messages or resisting reporting weight data out of indignation. Some "muted all notifications" from the program or blocked the WeChat groups "to concentrate on work." However, many forgot to unmute the notifications and gradually disengaged from their groups. Once participants left the virtual panopticon, surveillance lost its influence.

Table 2. Supporting quotes for theme 2: Cooperation and resistance to surveillance.

Subtheme	Quotes
Self-awareness	Quote 12: "As I had tried various ways to lose weight by myself using all kinds of methods and still couldn't control myself, I knew it would be better with someone watching you. Human beings are like that. When you're under supervision, you'll be mindful of what you're doing."
	Quote 13: "I felt I had to monitor myself. Self-management mattered because others were watching."
Untruthful data	Quote 14: "I needed to write lower numbers. Otherwise, the doctors would blame me."
	Quote 15: "When I didn't lose weight for 1 week, I felt embarrassed. I reported a figure and thought that I'd achieve it the next week, but I didn't. Then, the entire process was repeated the following week."
	Quote 16: "What we saw for a period of time was that everyone's weight records were about the same as they had always been, with no increase or decrease, when everyone was actually relapsing to some degree."
Invasive technology	Quote 17: "It seems that they're reckless with the notification times. For example, at 6 or 7 AM, when it is still early in the morning, my mobile phone beside my pillow would be bombarded with WeChat messages. 'Bong, Bong, Bong, Bong,' that's quite annoying."
	Quote 18: "Many times, those messages were sent when I was working. My work involves frequent conversations with my managers. It is unacceptable that my phone keeps buzzing."

Mutual Reinforcement of Peer Influences

Peer support, peer comparison, and surveillance had mutually reinforcing relationships that motivated participant engagement and health practices. Illustrative quotes for this theme are shown in Table 3. Peer support was mostly presented in the form of emotional support. Participants encountered similar inconveniences in their daily lives. Thus, they empathized with each other's feelings and experiences (quote 19). Joining a group with others who shared a similarity created a sense of belonging that engendered peer support (quote 20). Empathy, companionship, mutual encouragement, and collective empowerment were intangible resources shared by participants that bridged their feelings of isolation. The mantra of "solidarity is strength" also fueled their motivation to change (quote 21).

Table 3. Supporting quotes for theme 3: Mutual reinforcement of peer influences.

Subtheme	Quotes
Sense of belonging	Quote 19: "Since we were all fat, we could really understand what we felt deep down. Our families, friends, and colleagues all have some kind of prejudice against us because we were fat. In a clothes store, I often get neglected. We all know that feeling."
	Quote 20: "At home, I am the only person who is fat. Here, there are so many flabby people like me, so I don't feel as bad about myself. It gives me some sort of encouragement. I am really lazy, but when I see others trying hard to lose weight, I feel like I cannot allow myself to be left behind."
	Quote 21: "Like playing games, we are a team, and having a team is powerful. Just think about this, between one against five and five against one, which one is more powerful? Of course, it should be five against one."
Participatory surveillance	Quote 22: "Sharing the data demonstrates what you have done to lose weight. If you only lose a little, you will feel stressed. I would be envious of those who had made good progress when I had not, and I wondered whether I could do the same as them the next time."
	Quote 23: "I'd look at the ranking. There were some who had made progress but later they relapsed. I was glad to see that I relapsed just a little bit. I was kind of hoping others to do less well than me."
	Quite 24: "After all, it is what I want to do. Now I'm in this group, with others closely watching my progress, but I still think this depends on me, and it has little to do with others."
	Quote 25: "It is true that if I saw the records of others I would ask myself, 'Why did I rank last?' Even though I thought I should exercise that day, I would forget."
Negative group norms	Quote 26: "Everyone was striving to reach the goal. But this is not an easy course. So, there were some who wanted to have a rest, and this led others to think 'well, it might be fine if I also have a rest.'"
	Quote 27: "Someone would eat a little more and the rest would follow. It was inevitable that people would mutually influence each other in the group. I am the kind of person who could easily be influenced by others. Therefore, I also became less active."
Limited topics	Quote 28: "In the early days, I would talk about what I ate, how I exercised, what effects I saw, and how I overcame difficulties. Later, everyone knew how to tackle the problems and which environmental factors caused the failure to lose weight. It reached a point where nobody could help you with anything. For example, you're thinking about changing jobs because work-related stress was too intense for you to do anything else. Nobody could make that decision for you."
	Quote 29: "We're not really friends and thus we are shy when greeting each other. You know when people interact, we do not always go straight to the topic. When I tried to make a joke and only I responded, I would feel a bit awkward. If I wanted to raise a question about weight management, it was supposed to be the medical team who replied to me… When not talking about obesity theories, I felt that I was not familiar with anyone and should not post anything."

Participatory surveillance and peer comparison echoed one another in WeChat communication. While participants were the subjects of others' gaze, they also watched others self-evaluate. Seeing the progress of others drove positive competition and vicarious learning. All participants compared themselves to other members. They either speculated on others' effectiveness in weight loss or were secretly glad after finding that they were not the weakest in the group (quotes 22-23). Insofar as peer comparison was prevalent, participants noted that weight loss was a personal journey. They ultimately focused on their strength or lack of strength to sustain self-regulated behaviors rather than on their rankings in the group (quotes 24-25).

The interview data revealed that peer interactions were a double-edged sword. Positive forces derived from the interplay of peer comparison, peer support, and mutual supervision could be motivating. However, negative group dynamics may also occur to reshape group norms and collective practices. Participants noted that weight management was a tiring and difficult task. It was unavoidable that group members influenced each other (quotes 26-27). When group interactions moved toward a negative direction, the practice of mutual supervision and comparison reinforced the perception that "everyone does it; therefore, it is okay for me, too." Negative group norms, in turn, weakened engagement and retention.

Negative group dynamics could also result from the virtual community's limited capacity to enhance interpersonal bonds. Despite similar weight concerns, group members differed in age, occupation, work life, and medical symptoms. Over time, the few topics around weight management generated repetitive storytelling and information exchange that were not particularly useful in solving the root problems. The saturation of resource sharing explained many group members' regression from being open and expressive to being quiet and withdrawn (quote 28).

Finally, participants' preference for seeking information from medical professionals formed another barrier to developing deeper bonds among peers. When participants raised a question about weight management, they expected the medical team to reply to them. When not talking about obesity-related issues, they felt that they were not familiar with anyone and were hesitant to share their thoughts and perspectives (quote 29). The issue-specific communication style and participants' preferences for professional information sources narrowed the scope and depth of peer-to-peer communication. Peer interactions tended to dwindle after the initial excitement. Ultimately, long-term peer influences were primarily through the provision of reference points and superficial companionship that did not require deeper interpersonal ties.

In sum, our findings revealed that the cybersocial influences of social media interactions, presented in the forms of support, comparison, and surveillance, could be both positive and negative, and may change over time. Table 4 summarizes the possible influences.

Table 4. Social influences of social media interactions.

Dynamics	Doctor-patient communication	Peer-to-peer communication
Support		
Positive	Credible information sources	Sense of belonging
	Uncertainty reduction and confidence enhancement	Empathy
	Availability of timely feedback	Companionship and teamwork
	Enhanced trust and bonds	
Negative	Excuse for procrastination	Obstacles related to within-group heterogeneity
	Gap between knowing and doing	Limited topics for conversations
	Centralized communication	Unhelpful repetition of experience sharing
Surveillance		
Positive	Explicit request for adherence	Exhibition spaces of social information
	Enhanced vigilance for weight loss norms	Source of vicarious learning
	Basis for support-giving	Increased vigilance with disciplinary control
Negative	Lack of validation mechanism	Reinforcement of negative group norms
	Resistance to invasive technology	Collective inactivity
	Limited binding force	
Comparison		
Positive	Not applicable	Positive competition
		Enhanced view of oneself
		Realistic reference points
		Drive for self-improvement
Negative	Not applicable	Downturns in reference standards
		Collective tolerance for relapse
		Withdrawal after repeated failure in competition

Discussion

Principal Findings

WeChat facilitated the medical professionals' provision of information and tailored feedback that strengthened program participants' capacity to manage weight based on scientific knowledge. The panoptic gaze from medical professionals and peers reminded participants about the need to conform to weight control norms. However, the lack of validation mechanisms and participant resistance to invasive technology also limited the binding force of surveillance. Intragroup similarities derived peer support, encouraged participatory surveillance, and increased comparison, which are all deemed beneficial to engagement. However, peer influences were also jeopardized by negative group dynamics and the limited scope for peer-to-peer communication.

Comparison With Prior Work

Resonating with extant literature on social support, this study demonstrates that professional support and peer support perform different functions, and their modes of influence vary [14]. Professionally led support consisted of informational and appraisal support that were perceived to be prescriptive, directive, and affirmative. The approachability of medical professionals yielded an additional spillover effect of emotional support. Peer support mainly reflected emotional support that was mutual, noncompulsory, and empathic. The presence of medical professionals in support groups circumvented the problems of misinformation, tortuous discussions, and malicious interactions that occurred in interventions incorporating peer-to-peer communication only [12]. Professional support also reduced doubts about the perceived usefulness of support identified in previous studies [42]. However, support groups inclusive of medical professionals may invite stronger top-down control that could circumscribe the development of a vibrant online community and risk directing communication back to the conventional one-to-many model. Our results suggest that the inclusion of both professional and peer support in social media weight management interventions may diversify the provision of aid to fulfill participants' varying support needs. Nevertheless, a fine balance should be sought between the vertical communication of professional support and the horizontal communication of peer support to benefit from social media's multidimensionality, and to create a support network space that is both informative and participative.

Our qualitative findings offer possible explanations for the nonsignificant long-term effects of social support and the declining supportive communication over time in web-based weight management interventions found in previous studies [10,17]. Upon joining a new social media group, the knowledge gain and team support lacking in participants' previous experiences can enhance their efficacy and motivations through strengthened supervision and encouragement, and subsequently improve their self-regulated behaviors. However, the usefulness of professional support and peer support may reach saturation when there is no room to improve participants' knowledge level further and expand their online support networks. In this study, we specified the positive influences of support at the intellectual, psychological, and emotional levels, and we also illuminated the nuances underpinning the progressively smaller impacts of support over time. The findings of possible support saturation points, the use of knowledge gain as an excuse to delay weight loss efforts, and the variable directions of group dynamics in the evolving process of supportive communication add new information to the current understanding of support effects found in experimental studies [43-45].

This study reaffirmed the prevalence of peer comparison in cyberspace [20,23], and identified the mutual reinforcement of peer comparison, peer support, and surveillance in shaping group norms and engagement. Our findings support the notion that social media create an exhibition space that permits users to play multiple roles as supporters, comparison targets, and surveillants during group interactions [46]. In this study, the display of dietary records, weight data, and conversations increased intimacy among group members, and promoted comparison and peer-based surveillance. A supportive environment encouraged positive competition, which induced mutual encouragement and empathy. Likewise, a unilateral surveillance environment intensified peer comparison that, in turn, increased participant vigilance with mutual supervision. Similar to the few studies that examined the comparison component of social media weight management interventions [21,22,24], our results support the positive functions of peer comparison in enhancing motivations and friendly competition. However, the results also challenge a linear assumption about the positive relationship between peer comparison and weight behaviors. Group interactions underpinned by peer support, comparison, and surveillance can move toward both positive and negative directions that encourage or hinder engagement [24]. Collective tolerance for relapse, downturns in reference standards, and the tendency to withdraw after repeated failures of meeting group norms were consequences related to negative group dynamics. These results demonstrate the persuasive potential of peer influences in long-term communication, and suggest a need to attend to group norm changes in the design and moderation of social media weight management interventions.

In line with recent theoretical arguments about surveillance technology [26,30], this study identified two sources of panoptic gaze in social media. Together, public surveillance from medical professionals and participatory surveillance from peers created a reward and punishment system that reinforced self-awareness and self-monitoring in cyberspace. Those who conformed to disciplinary rules and achieved positive weight outcomes were rewarded when presenting positive figures in front of a virtual public. In contrast, punishments included displaying unsatisfactory weight outcomes and bearing the feelings of shame and guilt. Also aligned with past studies was the finding that power appeared in both surveillants' gaze and observees' responses to surveillance [29,47]. Although surveillance exerted a binding force on self-monitoring, participants could choose to stay in or exit from the virtual panopticon. Accordingly, there were tensions between the program's intent to maximize user vigilance with weight control norms through panoptic surveillance and participants' use of defensive strategies such as faking weight data and turning off notifications to reduce

burden and escape from the gaze. The various types of resistance are a complex result of participants' dissatisfaction with weight outcomes and their navigation between protecting a personal boundary and trading privacy for services. The permeation of participatory surveillance and its resonance with peer support and peer comparison may escalate online peer influences in unprecedented ways in the digital era. Nevertheless, surveillance technology also entails new issues such as the ease of data fabrication and privacy concerns that put new impediments on participant engagement. These are all worthy of further exploration and consideration.

Practical Implications

Our findings demonstrate the importance of sustaining positive group dynamics to prolong engagement and the necessity of developing preemptive measures against resistance. Organizing more homogenous groups based on sociodemographics and health concerns (eg, pregnancy) might help group members find more commonality of interests, and deepen their interpersonal bonds that encourage support sharing, friendly competition, and mutual supervision. The level of homophily may be better controlled and tested through improved design [48]. Adding new stimuli in different phases of an intervention, and changing the duration of each phase, might help fuel participant motivations and sustain positive group dynamics. Both are worth investigating in future trials. Improving the briefing of communication time points during the opening workshop might help attenuate resistance. Finally, having face-to-face communication with physicians and peers—and witnessing actual body shape changes—would provide irreplaceable validation and aspirational motivation. Identifying innovative ways to include offline interaction components such as periodic gatherings may be considered in the future.

Limitations

This study has three limitations. First, the qualitative nature of the study limits the generalization of the findings and the potential to test for associations among factors. Developing quantitative measures in experimental design can be the next step to verify the initial findings of the interplay of support, comparison, and surveillance revealed in this study. Second, it is unclear whether there is a difference between those who participated in the interviews and those who did not. In this study, we used maximum variation sampling to recruit men and women, young and old, and members from older and newer groups. Additional efforts were made to reach those who regained weight and those who withdrew from the WeChat groups. The diversity in the sample helped build a clear picture of social media interactions based on a wide range of experiences and observations. These actions may tackle some selection biases but cannot fully rule them out. Gaining more empirical support from future research may help validate the patterns observed in this study. Finally, social media–derived influences intervene in one's weight control behaviors to some extent at best. Weight management as an enduring process requires the balance of an array of intrapersonal, interpersonal, and environmental factors. This study pointed out a few areas for improvement that may help enhance the positive impacts of social media influences in weight management interventions. Nevertheless, interpretations of the results should be based on a realistic view of the limitations of cybersocial influences.

Conclusions

By inquiring into participant experiences using a weight management program for up to 2 years, this study reveals three aspects of social media influences derived from doctor-patient and peer-to-peer communication in cyberspace. Support, comparison, and surveillance play interrelated and fluctuant roles in motivating those with overweight or obesity to manage weight. When the dynamics are positive, social media communication strengthens disciplinary norms, encourages positive competition, and generates a binding force for engagement and retention. When the dynamics are negative, social media communication delays action, increases attrition, evokes resistance, and foments disengagement. From a theoretical perspective, the relationships among support, comparison, and surveillance identified in this study expand previously known aspects of social media influences, and advance understanding of the underlying processes of social media weight management interventions. From a practical perspective, this study offers empirical evidence that has implications for program improvement. Results of this study constitute a building block for the continuous advancement of employing social media technology to improve weight management and health.

Acknowledgments

This study was supported by the Commonweal Research Project on Application Technology in Zhejiang Province, China (Grant No. 2017C35006. Title: Research on modes of joint multidisciplinary interventions of obese population and prevention of metabolic diseases); Natural Science Foundation of Ningbo Municipality, China (Grant No. 2016A610169. Title: Research on modes of joint multidisciplinary interventions of obese population and prevention of metabolic diseases); China Healthcare Grant, Institute of Asia and Pacific Studies, University of Nottingham Ningbo China; and FHSS Competitive Research Funding, University of Nottingham Ningbo China. The authors would like to thank Liu Jia Ning and Yang Qian for their assistance with this project.

Authors' Contributions

LL oversaw the project and recruited participants. LC performed data collection, data analysis, and drafted the paper with inputs from KC, JL, MX, and LL. All authors contributed to the revision of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary of participant demographics.
[DOC File , 137 KB - mhealth_v9i3e19239_app1.doc]

References

1. Prevalence of overweight among adults, BMI ≥ 25, age-standardized estimates by country. Global Health observatory data repository.: World Health Organization; 2017 Sep 28. URL: http://apps.who.int/gho/data/view.main.CTRY2430A?lang=en [accessed 2021-02-18]
2. NCD Risk Factor Collaboration (NCD-RisC). Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19·2 million participants. Lancet 2016 Apr 02;387(10026):1377-1396 [FREE Full text] [doi: 10.1016/S0140-6736(16)30054-X] [Medline: 27115820]
3. Zhou L, Zeng Q, Jin S, Cheng G. The impact of changes in dietary knowledge on adult overweight and obesity in China. PLoS One 2017;12(6):e0179551 [FREE Full text] [doi: 10.1371/journal.pone.0179551] [Medline: 28644891]
4. Jensen MD, Ryan DH, Apovian CM, Ard JD, Comuzzie AG, Donato KA, American College of Cardiology/American Heart Association Task Force on Practice Guidelines, Obesity Society. 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and The Obesity Society. J Am Coll Cardiol 2014 Jul 01;63(25 Pt B):2985-3023 [FREE Full text] [doi: 10.1016/j.jacc.2013.11.004] [Medline: 24239920]
5. Bardus M, Smith JR, Samaha L, Abraham C. Mobile phone and web 2.0 technologies for weight management: a systematic scoping review. J Med Internet Res 2015;17(11):e259 [FREE Full text] [doi: 10.2196/jmir.5129] [Medline: 26573984]
6. Dumas A, Lapointe A, Desroches S. Users, uses, and effects of social media in dietetic practice: scoping review of the quantitative and qualitative evidence. J Med Internet Res 2018 Feb 20;20(2):e55 [FREE Full text] [doi: 10.2196/jmir.9230] [Medline: 29463487]
7. Jane M, Hagger M, Foster J, Ho S, Pal S. Social media for health promotion and weight management: a critical debate. BMC Public Health 2018 Jul 28;18(1):932 [FREE Full text] [doi: 10.1186/s12889-018-5837-3] [Medline: 30055592]
8. Waring ME, Jake-Schoffman DE, Holovatska MM, Mejia C, Williams JC, Pagoto SL. Social media and obesity in adults: a review of recent research and future directions. Curr Diab Rep 2018 Apr 18;18(6):34. [doi: 10.1007/s11892-018-1001-9] [Medline: 29671135]
9. Dahl AA, Hales SB, Turner-McGrievy GM. Integrating social media into weight loss interventions. Curr Opin Psychol 2016 Jun;9:11-15. [doi: 10.1016/j.copsyc.2015.09.018]
10. Chang T, Chopra V, Zhang C, Woolford SJ. The role of social media in online weight management: systematic review. J Med Internet Res 2013 Nov 28;15(11):e262 [FREE Full text] [doi: 10.2196/jmir.2852] [Medline: 24287455]
11. Williams G, Hamm MP, Shulhan J, Vandermeer B, Hartling L. Social media interventions for diet and exercise behaviours: a systematic review and meta-analysis of randomised controlled trials. BMJ Open 2014 Feb 12;4(2):e003926 [FREE Full text] [doi: 10.1136/bmjopen-2013-003926] [Medline: 24525388]
12. Wright KB, Bell SB, Wright KB, Bell SB. Health-related support groups on the internet: linking empirical findings to social support and computer-mediated communication theory. J Health Psychol 2003 Jan;8(1):39-54. [doi: 10.1177/1359105303008001429] [Medline: 22113899]
13. Goldsmith DJ, Albrecht TL. Social support, social networks, and health. In: Thompson TL, Parrott R, Nussbaum JF, editors. The Routledge handbook of health communication. New York: Routledge; 2011:335-348.
14. Hogan BE, Linden W, Najarian B. Social support interventions: do they work? Clin Psychol Rev 2002 Apr;22(3):383-442. [doi: 10.1016/s0272-7358(01)00102-7] [Medline: 17201192]
15. Gottlieb BH, Wachala ED. Cancer support groups: a critical review of empirical studies. Psychooncology 2007 May;16(5):379-400. [doi: 10.1002/pon.1078] [Medline: 16986205]
16. Kato-Lin Y, Padman R, Downs J, Abhishek V. Evaluating consumer m-Health services for promoting healthy eating: a randomized field experiment. AMIA Annu Symp Proc 2015;2015:1947-1956 [FREE Full text] [Medline: 26958294]
17. Hageman PA, Pullen CH, Hertzog M, Pozehl B, Eisenhauer C, Boeckner LS. Web-based interventions alone or supplemented with peer-led support or professional email counseling for weight loss and weight maintenance in women from rural communities: results of a clinical trial. J Obes 2017;2017:1602627. [doi: 10.1155/2017/1602627] [Medline: 28480078]
18. Fukuoka Y, Kamitani E, Bonnet K, Lindgren T. Real-time social support through a mobile virtual community to improve healthy behavior in overweight and sedentary adults: a focus group analysis. J Med Internet Res 2011 Jul 14;13(3):e49 [FREE Full text] [doi: 10.2196/jmir.1770] [Medline: 21752785]
19. Festinger L. A theory of social comparison processes. Human Relat 2016 Apr 22;7(2):117-140. [doi: 10.1177/001872675400700202]

20. Holland G, Tiggemann M. A systematic review of the impact of the use of social networking sites on body image and disordered eating outcomes. Body Image 2016 Jun;17:100-110. [doi: 10.1016/j.bodyim.2016.02.008] [Medline: 26995158]
21. Foster D, Linehan C, Kirman B, Lawson S, James G. Motivating physical activity at work: using persuasive social media for competitive step counting. 2010 Oct Presented at: MindTrek '10: Proceedings of the 14th International Academic MindTrek Conference: Envisioning Future Media Environments; October 6-8, 2010; Tampere, Finland p. 111-116. [doi: 10.1145/1930488.1930510]
22. Leahey TM, Kumar R, Weinberg BM, Wing RR. Teammates and social influence affect weight loss outcomes in a team-based weight loss competition. Obesity (Silver Spring) 2012 Jul;20(7):1413-1418. [doi: 10.1038/oby.2012.18] [Medline: 22310234]
23. Arigo D, Brown MM, Pasko K, Suls J. Social comparison features in physical activity promotion apps: scoping meta-review. J Med Internet Res 2020 Mar 27;22(3):e15642 [FREE Full text] [doi: 10.2196/15642] [Medline: 32217499]
24. Tong HL, Coiera E, Laranjo L. Using a mobile social networking app to promote physical activity: a qualitative study of users' perspectives. J Med Internet Res 2018 Dec 21;20(12):e11439 [FREE Full text] [doi: 10.2196/11439] [Medline: 30578201]
25. Lyon D. Surveillance studies: an overview. Cambridge, UK: Polity; 2007.
26. Lupton D. M-health and health promotion: the digital cyborg and surveillance society. Soc Theory Health 2012 Jun 27;10(3):229-244. [doi: 10.1057/sth.2012.6]
27. Depper A, Howe PD. Are we fit yet? English adolescent girls' experiences of health and fitness apps. Health Sociology Review 2016 Jul 04;26(1):98-112. [doi: 10.1080/14461242.2016.1196599]
28. Foucault M. Discipline and punish: the birth of the prison. London, UK: Penguin; 1977.
29. Albrechtslund A. Online social networking as participatory surveillance. First Monday 2008 Mar 03;13(3). [doi: 10.5210/fm.v13i3.2142]
30. Marwick A. The public domain: surveillance in everyday life. Surveill Soc 2012 Jun 20;9(4):378-393. [doi: 10.24908/ss.v9i4.4342]
31. Hawn C. Take two aspirin and tweet me in the morning: how Twitter, Facebook, and other social media are reshaping health care. Health Aff (Millwood) 2009;28(2):361-368. [doi: 10.1377/hlthaff.28.2.361] [Medline: 19275991]
32. Goodyear VA, Kerner C, Quennerstedt M. Young people's uses of wearable healthy lifestyle technologies; surveillance, self-surveillance and resistance. Sport Educ Soc 2019;24(3):212-225. [doi: 10.1080/13573322.2017.1375907]
33. eMarketer. WeChat user penetration in China, by city tier, Sept 2015. INSIDER Intelligence. 2015 Dec 01. URL: http://www.emarketer.com/Chart/WeChat-User-Penetration-China-by-City-Tier-Sep-2015-of-population-each-group/181854 [accessed 2019-06-27]
34. eMarketer. WeChat users and penetration in China, 2016-2021. INSIDER Intelligence. 2017 Jul 01. URL: https://tinyurl.com/87nbe5nr [accessed 2019-06-27]
35. Messier SP, Loeser RF, Miller GD, Morgan TM, Rejeski WJ, Sevick MA, et al. Exercise and dietary weight loss in overweight and obese older adults with knee osteoarthritis: the Arthritis, Diet, and Activity Promotion Trial. Arthritis Rheum 2004 May;50(5):1501-1510. [doi: 10.1002/art.20256] [Medline: 15146420]
36. Xu M, Chattopadhyay K, Li J, Rai N, Chen Y, Hu F, et al. Weight management programme for overweight and obese adults in Ningbo, China: a feasibility pre- and post-intervention study. Front Public Health 2019;7:388. [doi: 10.3389/fpubh.2019.00388] [Medline: 31921752]
37. Suri H. Purposeful sampling in qualitative research synthesis. Qual Res J 2011 Aug 03;11(2):63-75. [doi: 10.3316/qrj1102063]
38. Strauss A, Corbin J. Basics of qualitative research techniques. Thousand Oaks, CA: Sage; 1998.
39. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
40. Whittemore R, Chase SK, Mandle CL. Validity in qualitative research. Qual Health Res 2001 Jul;11(4):522-537. [doi: 10.1177/104973201129119299] [Medline: 11521609]
41. Koelsch LE. Reconceptualizing the Member Check interview. Int J Qual Method 2013 Feb;12(1):168-179. [doi: 10.1177/160940691301200105]
42. Pagoto S, Waring ME, May CN, Ding EY, Kunz WH, Hayes R, et al. Adapting behavioral interventions for social media delivery. J Med Internet Res 2016 Jan 29;18(1):e24 [FREE Full text] [doi: 10.2196/jmir.5086] [Medline: 26825969]
43. Greaves CJ, Sheppard KE, Abraham C, Hardeman W, Roden M, Evans PH, IMAGE Study Group. Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions. BMC Public Health 2011 Feb 18;11:119 [FREE Full text] [doi: 10.1186/1471-2458-11-119] [Medline: 21333011]
44. Olander EK, Fletcher H, Williams S, Atkinson L, Turner A, French DP. What are the most effective techniques in changing obese individuals' physical activity self-efficacy and behaviour: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2013 Mar 03;10:29 [FREE Full text] [doi: 10.1186/1479-5868-10-29] [Medline: 23452345]
45. Meads C, Exley J. A systematic review of group walking in physically healthy people to promote physical activity. Int J Technol Assess Health Care 2018 Jan 17;34(1):27-37. [doi: 10.1017/s0266462317001088]
46. Hogan B. The presentation of self in the age of social media: distinguishing performances and exhibitions online. Bull Sci Technol Soc 2010 Nov 01;30(6):377-386. [doi: 10.1177/0270467610385893]

47. Best K. Living in the control society: surveillance, users and digital screen technologies. Int J Cultur Stud 2010 Jan 06;13(1):5-24. [doi: 10.1177/1367877909348536]
48. Zhang S, de la Haye K, Ji M, An R. Applications of social network analysis to obesity: a systematic review. Obes Rev 2018 Jul;19(7):976-988. [doi: 10.1111/obr.12684] [Medline: 29676508]

Edited by L Buis; submitted 09.04.20; peer-reviewed by D Frohlich, A McKinlay, S Hoffman Marchese; comments to author 01.07.20; revised version received 24.08.20; accepted 01.02.21; published 01.03.21.

Please cite as:
Chang L, Chattopadhyay K, Li J, Xu M, Li L
Interplay of Support, Comparison, and Surveillance in Social Media Weight Management Interventions: Qualitative Study
JMIR Mhealth Uhealth 2021;9(3):e19239
URL: https://mhealth.jmir.org/2021/3/e19239
doi:10.2196/19239
PMID:33646130

Original Paper

Effectiveness of the Alfalfa App in Warfarin Therapy Management for Patients Undergoing Venous Thrombosis Prevention and Treatment: Cohort Study

Hua Cao[1,2], PhD; Shaojun Jiang[3,4], BS; Meina Lv[3,4], BS; Tingting Wu[3,4], BS; Wenjun Chen[3,4], BS; Jinhua Zhang[3,4], PhD

[1]Department of Cardiac Surgery, Fujian Medical University Union Hospital, Fuzhou, China

[2]Department of Cardiac Surgery, Fujian Maternity and Children Health Hospital, Fuzhou, China

[3]Department of Pharmacy, Fujian Medical University Union Hospital, Fuzhou, China

[4]College of Pharmacy, Fujian Medical University, Fuzhou, China

Corresponding Author:
Jinhua Zhang, PhD
Department of Pharmacy
Fujian Medical University Union Hospital
#29 Xinquan Road, China
Fuzhou, 350001
China
Phone: 86 591 83357896 ext 8301
Fax: 86 59183357896
Email: pollyzhang2006@126.com

Abstract

Background: Over the years, the internet has enabled considerable progress in the management of chronic diseases, especially hypertension and diabetes. It also provides novel opportunities in online anticoagulation management. Nevertheless, there is insufficient evidence regarding the effectiveness of online anticoagulation management.

Objective: This study explored the effectiveness and safety of warfarin management via the Alfalfa app, so as to provide evidence in support of anticoagulant management through online services.

Methods: In this retrospective, observational cohort study, 824 patients were included. In the offline group, patients went to the hospital clinic for warfarin management. In the Alfalfa app group, patients reported the dose of warfarin, current international normalized ratio (INR) value, and other related information through the Alfalfa app. Physicians or pharmacists used the app to adjust the dose of warfarin and determined the time for the next blood INR testing. Patients completed INR testing by point-of-care at home or hospital. The primary outcome of the study was the percentage of time in therapeutic range (TTR). Secondary outcomes included minor and major bleeding events, thrombotic events, warfarin-related emergency department visits, hospital admissions, and high INR values.

Results: The TTR and percentage of INR values in the range were significantly higher in the Alfalfa app group than in the offline group (79.35% vs 52.38%, P<.001; 3314/4282, 77.39% vs 2005/4202, 47.72%, P<.001, respectively). Patients managed via the Alfalfa app had lower rates of subtherapeutic (172/4282, 4.02% vs 388/4202, 9.23%; P<.001), supratherapeutic (487/4282, 11.37% vs 882/4202, 20.99%; P<.001), and extreme subtherapeutic INR values (290/4282, 6.77% vs 910/4202, 21.66%; P<.001). Additionally, the Alfalfa app group had lower incidences of major bleeding (2/425, 0.5% vs 12/399, 3.0%; P=.005), warfarin-related emergency department visits (13/425, 3.1% vs 37/399, 9.3%; P<.001), and hospital admissions (1/425, 0.2% vs 12/399, 3.0%; P=.001) compared with the offline group. However, the Alfalfa app group had a higher incidence of minor bleeding than the offline group (45/425, 10.6% vs 20/399, 5.0%; P=.003). There were similar incidences in extreme supratherapeutic INR values (19/4282, 0.44% vs 17/4202, 0.40%; P=.78) and thromboembolic events (1/425, 0.2% vs 1/399, 0.3%; P=.53) between the two groups.

Conclusions: Warfarin management is superior via the Alfalfa app than via offline services in terms of major bleeding events, warfarin-related emergency department visits, and hospital admissions.

(JMIR Mhealth Uhealth 2021;9(3):e23332) doi:10.2196/23332

KEYWORDS

warfarin; anticoagulation; smartphone; telemedicine; app; online; offline; bleeding; INR; TTR; mobile phone

Introduction

Venous thrombosis, which includes deep vein thrombosis, pulmonary embolism, and cardiogenic stroke, poses a significant burden on the health care system [1]. Anticoagulants are effective in the treatment of venous thrombosis. Patients with atrial fibrillation and valve replacement surgery are at high risk for venous thrombosis; therefore, these patients require anticoagulant treatment. Warfarin is the most commonly used oral anticoagulant due to its low price, specific antagonists, and substantial evidence-based clinical data [2]. The dose of warfarin required for anticoagulation is closely related to dietary vitamin K intake, concurrent medications, the presence of other diseases, body weight, and aerobic exercise. Therefore, warfarin doses during anticoagulation therapy need to be adjusted according to the results of international normalized ratio (INR) tests. Without proper adjustment, excessive warfarin may cause bleeding, and inadequate warfarin doses may cause thrombosis and possible death [3]. This represents a costly use of medical resources and is physically and psychologically traumatic to patients and their families [4].

Warfarin dose adjustment is usually performed at hospital clinics. For patients living in rural areas, adjustment of warfarin dosages requires transportation, accommodation, and time. Patients with low incomes are willing to risk thrombosis and bleeding in an attempt to reduce costs. Online services such as smartphones, text messages, Bluetooth, and communication platforms may assist in adjusting warfarin doses [4-6]. Our meta-analysis, which included 16,915 patients, of whom 8655 had their warfarin dosage adjusted in the hospital and 8260 had their warfarin dosages adjusted through online services, revealed that online services were associated with fewer warfarin-related hospital admissions than hospital management (odds ratio 0.47, 95% CI 0.30-0.73; $P<.001$). However, there was no statistically significant difference in other anticoagulant effectiveness or clinical outcomes between the two groups [4]. Our previous research provided online services for patients who were taking warfarin through QQ group communication platforms, which can be accessed via smartphones, computers, laptops, and tablets. QQ is an instant messaging software service developed by Chinese company Tencent Holdings Limited. As described by You et al, social groups are some of “the main features of QQ [that allow] multiple users to communicate instantly. A message posted by a member is immediately received by all the other group members” [7]. The results showed that online services yielded similar clinical outcomes to hospital services, even though the incidence of supratherapeutic INR values increased [6]. However, there were some limitations in the QQ group communication platform. First, one patient reported his medical condition, and all the other patients could see it. There was no method of protecting patient privacy. Second, it was challenging to obtain the most recent medical information of patients because the information on all patients was mixed. Third, the information could only exist temporarily.

Recently, we developed a new warfarin dosage adjustment app named Alfalfa, which is available for installation on a mobile device; the communication between the physicians or pharmacists and the patient is point-to-point. The privacy of the patients is well protected. Meanwhile, personal medical information can be retrieved by the patients or their physicians or pharmacists at any time. As far as we know, apart from Alfalfa, there are only two studies focusing on managing warfarin with the use of mobile apps that can automatically suggest a dose and time for the next blood test based on the patient's INR value [8,9]. The difference between these mobile phone apps and Alfalfa is that Alfalfa provides a channel for communication between patients and medical staff, while the other two warfarin management apps focus on patient self-management. Alfalfa consists of a background management system and a remote anticoagulation management system. This remote anticoagulation management system can be divided into a patient terminal called Alfalfa Health Management and a medical terminal called Alfalfa Anticoagulation Guidance. The main function of Alfalfa is to facilitate the exchange of information between patients and medical staff, so that patients can report blood coagulation results and changes in their health status to physicians or pharmacists through the internet and then take their medicine and check the INR value of warfarin according to the advice of physicians or pharmacists [10]. In this study, we explored the effectiveness and safety of warfarin dose management via the Alfalfa app, which can provide further evidence in support of anticoagulant management via online anticoagulation services.

Methods

Study Design and Participants

We conducted a retrospective, observational cohort study to explore the effectiveness and safety of warfarin management via the Alfalfa app versus offline warfarin management. Participants were enrolled between December 2016 and March 2019 in Fujian Medical University Union Hospital (FMUUH). Inclusion criteria were patients who (1) received warfarin therapy for at least 3 months, (2) were willing to learn and accept the Alfalfa app or offline warfarin management, and (3) were willing to undergo follow-up. Exclusion criteria were pregnancy; planning to change to other anticoagulants; serious bleeding, thrombotic events, or both in the previous 3 months; and adjustment of warfarin doses by the patients themselves without a physician’s order. The primary outcome of the study, that is, percentage of time in therapeutic range (TTR), was calculated using a standard linear interpolation method [11]. Secondary outcomes included minor bleeding events, major bleeding events, thrombotic events, warfarin-related emergency department visits, warfarin-related hospital admissions, and high INR values. Major bleeding events included any bleeding requiring hospitalization or transfusion, as defined in the International Society on Thrombosis and Haemostasis classification [12]. Minor bleeding events included nose bleeding, conjunctival bleeding, gum bleeding, subcutaneous

purpura, menstrual abnormalities (increased, prolonged, or advanced), and other bleeding symptoms that can be quickly stopped.

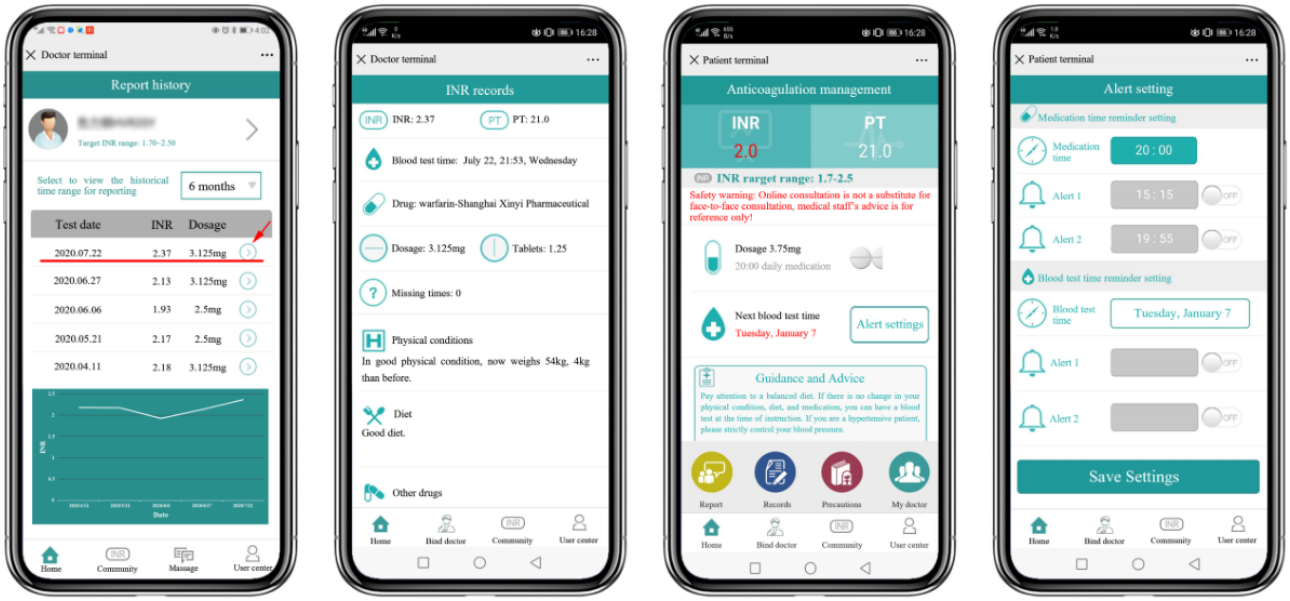
A warfarin teaching booklet was given to the patients for further reading prior to counseling as part of the standardized education process [13]. Then, the clinical pharmacist asked the patients which method of warfarin dosage adjustment would be selected after discharge: offline services or Alfalfa app. For offline services, patients could go to the local or big city hospital clinics for warfarin management. Patients were required to learn about warfarin through the booklet and complete a paper questionnaire (Multimedia Appendix 1) in wards or anticoagulation clinic. A validated questionnaire was chosen to determine the degree of anticoagulation knowledge [14]. For the Alfalfa app warfarin management, patients needed to download and install the app on their smartphones. Additionally, patients were instructed on how to use the app. Prior to registration, patients were required to learn about warfarin and complete the questionnaire. When a questionnaire score greater than or equal to 90 (out of 100) was achieved, they could register. The registration information includes name, age, body weight, anticoagulation indication, other diseases, concurrent medications, and dietary habits. Patients may select one of the physicians or pharmacists in the Alfalfa app to modify their warfarin dosage. The selected physician or pharmacist accepts the request from each patient. Patients may complete an INR test by point-of-care at home or hospital.

The protocols of warfarin dose adjustment are determined by the anticoagulation team in the FMUUH. Monitoring is frequently performed at the beginning of warfarin therapy, with at least one visit per week in the first two weeks of therapy, followed by one visit every two to four weeks. The target INR range is 1.7-2.5 for patients with valve replacement and 2.0-3.0 for those with atrial fibrillation and venous thrombosis. Our study protocol was approved by the Ethics Committee of FMUUH (2016KY036).

Data Collection

The data automatically recorded in the Alfalfa app included demographic characteristics, indication for warfarin therapy, duration of anticoagulation, target INR range, INR at each report, warfarin dose at each report, bleeding events, thromboembolic events, emergency department visits, and hospitalization (Figure 1A and B). Patients received alerts and reminders on subsequent INR testing and time for daily warfarin doses (Figure 1C and D). When the online appointment time arrived, patients completed the INR testing and reported their INR results, warfarin dosage, disease state, concurrent medications, and dietary habits. Then, they received information on the adjusted warfarin dosage, time for subsequent INR testing, and recommendations on diet and exercise from physicians or pharmacists. This information was automatically collected in the Alfalfa app. The data from the offline warfarin management service were collected from electronic medical records and telephone follow-ups.

Figure 1. Alfalfa app.



Statistical Analysis

Statistical analysis was performed using Microsoft Excel, version 2019 (Microsoft Corp), and SPSS, version 20.0 (IBM Corp).

For different variables in the baseline characteristics and results, different statistical methods were used to check the differences. An independent two-sample, two-tailed *t* test was used for statistical analysis of continuous variables, including age and TTR. A chi-square test or Fisher exact test was used for categorical variables, including sex, conditions, diseases, target INR, result of efficacy, adverse events, and INR results. Continuous variables are expressed as mean (SD), and categorical variables are expressed as quantity (percentage). Statistical significance is set at $P<.05$.

Results

Participant Characteristics

A total of 824 patients were enrolled and assigned to either the offline warfarin management group (n=399) or the Alfalfa app group (n=425). There were no significant differences in age, sex, conditions, diseases, coprescribed antiplatelet therapy, or target INR between the groups. In the current study, 48.07% of the participants were men (Table 1). The main condition was heart valve replacement, and the main disease was hypertension.

Table 1. Patient characteristics.

Characteristics	Offline (n=399)	Alfalfa app (n=425)	*P* value
Age (years), mean (SD)	53.00 (12.28)	51.41 (12.54)	.07
Male sex, n (%)	175 (43.86)	237 (55.76)	.23
Conditions, n (%)			
Heart valve replacement	358 (89.72)	381 (89.65)	.97
Atrial fibrillation	22 (5.51)	22 (5.18)	.83
Vein thromboembolism	19 (4.76)	22 (5.18)	.78
Diseases, n (%)			
Hypertension	73 (18.30)	60 (14.12)	.10
Hyperuricemia or gout	9 (2.26)	11 (2.59)	.76
Diabetes	15 (3.76)	15 (3.53)	.86
Co-prescribed antiplatelet therapy	7 (1.75)	8 (1.88)	.89
Target INR[a], n (%)			
1.5-2.0	107 (26.82)	126 (29.65)	.37
1.7-2.5	271 (67.92)	277 (65.18)	.40
2.0-3.0	21 (5.26)	22 (5.17)	.96

[a]INR: international normalized ratio.

Anticoagulant Control and Adverse Events

The TTR was significantly higher in the Alfalfa app group than in the offline warfarin management group (79.35% vs 52.38%, *P*<.001). Furthermore, the Alfalfa app group had a higher percentage of INR values in range than the offline group (3314/4282, 77.39% vs 2005/4202, 47.72%; *P*<.001). Patients managed by the Alfalfa app had lower rates of subtherapeutic (172/4282, 4.02% vs 388/4202, 9.23%; *P*<.001), supratherapeutic (487/4282, 11.37% vs 882/4202, 20.99%; *P*<.001), and extreme subtherapeutic INR values (290/4282, 6.77% vs 910, 21.66%; *P*<.001) compared to those managed by offline services. There were similar incidences in extreme supratherapeutic INR values (19/4282, 0.44% vs 17/4202, 0.40%; *P*=.78) between the two groups (Table 2).

Table 2. Anticoagulant control.

INR[a] value	Offline (n=4202)	Online (n=4282)	*P* value
Time in therapeutic range (%), mean (SD)	52.38 (12.67)	79.35 (26.31)	<.001
Extreme subtherapeutic, n (%)	910 (21.66)	290 (6.77)	<.001
Subtherapeutic, n (%)	388 (9.23)	172 (4.02)	<.001
Therapeutic, n (%)	2005 (47.72)	3314 (77.39)	<.001
Supratherapeutic, n (%)	882 (20.99)	487 (11.37)	<.001
Extreme supratherapeutic, n (%)	17 (0.40)	19 (0.44)	.78

[a]INR: international normalized ratio.

The incidences of major bleeding events (2/425, 0.5% vs 12/399, 3.0%; *P*=.005), warfarin-related emergency department visits (13/425, 3.1% vs 37/399, 9.3%; *P*<.001), and warfarin-related hospital admissions (1/425, 0.2% vs 12/399, 3.0%; *P*=.001) were lower in the Alfalfa app group than in the offline management group. However, the Alfalfa app group had a higher incidence of minor bleeding events than the offline group (45/425, 10.6% vs 20/399, 5.0%; *P*=.003). There were similar incidences in thromboembolic events (1/425, 0.2% vs 1/399, 0.3%; *P*=.53) between the two groups (Table 3).

Table 3. Adverse events.

Characteristics	Offline (n=399), n (%)	Alfalfa app (n=425), n (%)	*P* value
Major bleeding events	12 (3.1)	2 (0.5)	.005
Minor bleeding events	20 (5.0)	45 (10.6)	.003
Thromboembolic events	1 (0.3)	1 (0.2)	.53
Warfarin-related emergency department visits	37 (9.3)	13 (3.1)	<.001
Warfarin-related hospital admissions	12 (3.0)	1 (0.2)	.001

Discussion

Patients taking warfarin need to monitor INR and adjust their warfarin dosages accordingly. The included patients took warfarin mainly for the following three reasons: heart valve replacement, vein thromboembolism, and atrial fibrillation. Warfarin is the only oral anticoagulant available for thrombosis prevention after heart valve replacement [15]. The type of prosthetic valve, its anatomical location, and patient-specific risks of thromboembolism and bleeding influence the specific intensity and duration of antithrombotic treatment to prevent prosthetic valve thrombosis [16]. Some studies have proved that Chinese people need lower anticoagulation intensity INR (1.5-2.5) to warfarin in comparison to the recommended INR (2.5-3.5) in developed countries [17-19]. As long-term anticoagulant therapy for vein thromboembolism, an authoritative guideline has suggested warfarin adjusted to achieve an INR of 2.0-3.0 [20]. In patients with atrial fibrillation, warfarin reduces the relative risk of stroke by 64% and all-cause mortality by 26% [21]. Clinical studies have confirmed that when INR ranges between 2.0 and 3.0 in patients with atrial fibrillation, warfarin effectively prevents stroke and does not significantly increase the risk of bleeding [22]. The effectiveness of anticoagulant therapy is usually expressed as TTR or INR within the therapeutic target range. TTR values greater than 65% are indicative of effective anticoagulation therapy [21]. Due to the different medical standards and management methods in different regions, the TTR range is wide (29%-75%) [23]. Even in strictly controlled, large-scale clinical trials such as the ROCKET AF study, the average TTR was only 55.2% for all patients, among whom Chinese patients with atrial fibrillation had a TTR of only 47%. It was challenging for anticoagulation therapy to achieve the desired effect [23,24].

In this study, TTR was significantly higher in the Alfalfa app group than in the offline warfarin management group (79.35% vs 52.38%, $P<0.001$). These results were similar to those reported by Prochaska et al, who showed that TTR was significantly higher from telemedicine-based coagulation service than from regular anticoagulation management (75.5% vs 66.3%, $P<0.001$) [25]. However, a study by Lee et al revealed opposite results [26]. Some studies have reported similar TTR between online anticoagulation management and regular medical care [27-29]. This study showed that the rates of subtherapeutic, supratherapeutic, and extreme subtherapeutic INR values were significantly lower in the Alfalfa app group than in the offline warfarin management group. The lower rate of abnormal INR distribution may reduce the incidence of bleeding and thrombotic events.

Our research findings showed that the Alfalfa app had lower incidences of major bleeding events, warfarin-related emergency department visits, and warfarin-related hospital admissions than the offline warfarin management group. However, the incidence of minor bleeding events was higher in the Alfalfa app group. This finding was similar to the results reported by Blissit et al, who concluded that the incidence of minor bleeding events was higher in the online anticoagulation management group [30]. The top three minor bleeding types in the offline group were nose bleeding (7 cases), gum bleeding (7 cases), and hematuria (4 cases). The top three minor bleeding types in the Alfalfa app were gum bleeding (10 cases), menstrual abnormalities (increased, prolonged, or advanced; 7 cases), and subcutaneous purpura (9 cases). More minor bleeding events in the Alfalfa app may be due to the way patients provided feedback. The data of the offline group are obtained through follow-up, while the data of the Alfalfa app group are actively reported by the patient in the app. This study is a retrospective study. Most patients received follow-up calls after taking warfarin for 3 months. It is possible that patients could not recall the minor bleeding adverse reactions and relay them to the investigator. On the other hand, patients in the Alfalfa app group have a stronger willingness to communicate with medical staff. Thus, the Alfalfa app showed more minor bleeding events.

The studies by Xia et al and Cryder et al showed that there were fewer anticoagulation-related hospital admissions in the online anticoagulation group [4,27]. Most studies have reported that there are no differences in the incidence of major bleeding events between the two groups [5,27-32]. Our findings revealed that the incidences of thromboembolic events were similar between the two groups. While the offline anticoagulant management models were similar among the studies, the online management models were slightly different. For example, some studies managed patients through text messages, while others relied on Bluetooth and online communication platforms. The Alfalfa app is the first app specifically designed for warfarin dose adjustment management. This study showed that the application of the app not only achieved better anticoagulant control effects, but also greatly reduced major bleeding events, emergency department visits, and hospitalization. Therefore, the Alfalfa app is likely to reduce related medical costs. There are two main reasons for the positive effects of the Alfalfa app. First, the Alfalfa app can retrieve the patient's previous warfarin dosage, INR value, bleeding and thrombotic events, and adverse reaction history so that physicians and pharmacists can better analyze the reason for abnormal INR or bleeding. Physicians and pharmacists can assess whether the abnormal INR or bleeding events are caused by warfarin or induced by other

diseases. Second, the Alfalfa app has an automatic response function for the physicians or pharmacists according to the protocols of warfarin dose adjustment. After the automatic response, physicians or pharmacists can check whether the automatic response is correct before it is submitted to the patients. This function is considerably helpful for anticoagulation management newcomers.

However, the results of this study showed that anticoagulation control may not be completely consistent with the clinical outcomes. For example, the extreme supratherapeutic INR values in this study were distributed close to each other, but the incidence of major bleeding events was lower in the Alfalfa app group. The average TTR of the offline warfarin management group was only 52.38% (<60%). In theory, the incidence of thrombosis should be higher in the offline warfarin management group, but the incidences of thrombosis were similar between the two groups.

In recent years, the internet has made considerable progress in the management of chronic diseases, especially hypertension and diabetes. The internet provides novel opportunities in anticoagulation management. Telemedicine breaks geographical and spatial limitations, providing high-quality medical resources for patients living in vast rural and remote areas. Similarly, INR real-time detection technology is convenient for at-home INR detection. This technology allows patients to enjoy high-quality online anticoagulation management services at home.

Currently, rural and medically underresourced areas often lack specialist anticoagulant physicians or pharmacists, resulting in a significant risk of bleeding and thrombosis for patients living in these areas. We are conducting a national randomized controlled multicenter study comparing the effects of the Alfalfa app and offline services [33]. If the Alfalfa app becomes widely available, there could be significant benefits for such patients. At the same time, we also look forward to the economic evaluation results of these two management models.

There were some limitations to this study. First, as a retrospective study, subjects were not assigned to either group; therefore, selection bias may be present. Second, the participants were included from a single center in Southeast China, which may not be representative of all patients with anticoagulation therapy. More prospective, randomized, and multicenter studies are required to confirm our findings.

In conclusion, warfarin management in the Alfalfa app group was superior to that in the offline group in terms of TTR, abnormal INR, major bleeding events, warfarin-related emergency department visits, and warfarin-related hospital admissions. Warfarin management via the Alfalfa app may be suitable for patients living in rural and remote areas.

Acknowledgments

This study was funded by the Natural Science Foundation of Fujian Province of China (2018Y0037). The authors are grateful to the Alfalfa Health Management Company for the donation of their devices and to Siyi Cai, the study coordinator, for her dedication in ensuring data accuracy.

Authors' Contributions

This study was designed by JZ. HC and JZ wrote the manuscript. SJ performed the statistical analysis. ML, WC, and TW participated in the study and contributed to data analysis, drafting, data interpretation, and manuscript revision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire.
[DOCX File , 21 KB - mhealth_v9i3e23332_app1.docx]

References

1. Wolberg AS, Rosendaal FR, Weitz JI, Jaffer IH, Agnelli G, Baglin T, et al. Venous thrombosis. Nat Rev Dis Primers 2015 May 07;1:15006. [doi: 10.1038/nrdp.2015.6] [Medline: 27189130]
2. Zhang J, Wu T, Chen W, Fu J, Xia X, Chen L. Effect of Gene-Based Warfarin Dosing on Anticoagulation Control and Clinical Events in a Real-World Setting. Front Pharmacol 2019;10:1527 [FREE Full text] [doi: 10.3389/fphar.2019.01527] [Medline: 32038232]
3. Zhang J, Liu M, Chen Q, Wu J, Cao H. Outcomes of an online pharmacist-managed anticoagulation clinic for individuals on warfarin therapy living in rural communities. Thromb Res 2017 Sep;157:136-138. [doi: 10.1016/j.thromres.2017.07.016] [Medline: 28746903]
4. Xia X, Wu J, Zhang J. The effect of online versus hospital warfarin management on patient outcomes: a systematic review and meta-analysis. Int J Clin Pharm 2018 Dec;40(6):1420-1429. [doi: 10.1007/s11096-018-0731-8] [Medline: 30367374]
5. McCahon D, Baker JM, Murray ET, Fitzmaurice DA. Assessing the utility of an online registry for patients monitoring their own warfarin therapy. J Clin Pathol 2016 Apr;69(4):331-336. [doi: 10.1136/jclinpath-2015-203168] [Medline: 26519487]

6. Cao H, Wu J, Zhang J. Outcomes of warfarin therapy managed by pharmacists via hospital anticoagulation clinic versus online anticoagulation clinic. Int J Clin Pharm 2018 Oct;40(5):1072-1077. [doi: 10.1007/s11096-018-0674-0] [Medline: 29956134]
7. You ZQ, Han XP, Lü L, Yeung CH. Empirical Studies on the Network of Social Groups: The Case of Tencent QQ. PLoS One 2015;10(7):e0130538 [FREE Full text] [doi: 10.1371/journal.pone.0130538] [Medline: 26176850]
8. Li Y, Dong L, Xiang D, Zhang Y, Chen X, Long J, et al. Patient compliance with an anticoagulation management system based on a smartphone application. J Thromb Thrombolysis 2019 Aug;48(2):263-269. [doi: 10.1007/s11239-019-01859-0] [Medline: 31028513]
9. Smaradottir B, Martinez S, Borycki E, Loudon G, Kushniruk A, Jortveit J, et al. User Evaluation of a Smartphone Application for Anticoagulation Therapy. Stud Health Technol Inform 2018;247:466-470. [Medline: 29678004]
10. Jiang S, Lv M, Wu T, Wenjun C, Zhang J. The Development and Implementation of Alfalfa: A Smartphone Application for Remote Adjustment of Warfarin Dose. JMIR Preprints. Preprint posted online on October 8, 2020 2021. [doi: 10.2196/preprints.24625]
11. Rosendaal FR, Cannegieter SC, van der Meer FJM, Briët E. A method to determine the optimal intensity of oral anticoagulant therapy. Thromb Haemost 1993 Mar 01;69(3):236-239. [Medline: 8470047]
12. Schulman S, Kearon C, Subcommittee on Control of Anticoagulation of the Scientific and Standardization Committee of the International Society on Thrombosis and Haemostasis. Definition of major bleeding in clinical investigations of antihemostatic medicinal products in non-surgical patients. J Thromb Haemost 2005 Apr;3(4):692-694 [FREE Full text] [doi: 10.1111/j.1538-7836.2005.01204.x] [Medline: 15842354]
13. Cao H, Wu T, Chen W, Fu J, Xia X, Zhang J. The effect of warfarin knowledge on anticoagulation control among patients with heart valve replacement. Int J Clin Pharm 2020 Jun;42(3):861-870. [doi: 10.1007/s11096-020-01043-y] [Medline: 32410208]
14. Briggs AL, Jackson TR, Bruce S, Shapiro NL. The development and performance validation of a tool to assess patient anticoagulation knowledge. Res Social Adm Pharm 2005 Mar;1(1):40-59. [doi: 10.1016/j.sapharm.2004.12.002] [Medline: 17138465]
15. Nishimura R, Otto C, Bonow R, Carabello BA, Erwin JP, Fleisher LA, et al. 2017 AHA/ACC Focused Update of the 2014 AHA/ACC Guideline for the Management of Patients With Valvular Heart Disease: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation 2017 Jun 20;135(25):e1159-e1195. [doi: 10.1161/CIR.0000000000000503] [Medline: 28298458]
16. Dangas GD, Weitz JI, Giustino G, Makkar R, Mehran R. Prosthetic Heart Valve Thrombosis. J Am Coll Cardiol 2016 Dec 20;68(24):2670-2689 [FREE Full text] [doi: 10.1016/j.jacc.2016.09.958] [Medline: 27978952]
17. Tao H, Li Q, Zhou Q, Chen J, Fu B, Wang J, CLIATHVR multicenter clinical study team. A prediction study of warfarin individual stable dose after mechanical heart valve replacement: adaptive neural-fuzzy inference system prediction. BMC Surg 2018 Feb 15;18(1):10 [FREE Full text] [doi: 10.1186/s12893-018-0343-1] [Medline: 29448930]
18. Dong L, Shi Y, Xu J, Zhang E, Liu J, Li Y, et al. [The multicenter study on the registration and follow-up of low anticoagulation therapy for the heart valve operation in China]. Zhonghua Yi Xue Za Zhi 2016 May 24;96(19):1489-1494. [doi: 10.3760/cma.j.issn.0376-2491.2016.19.006] [Medline: 27266493]
19. Zhou X, Zhuang W, Hu J, Li J, Yu J, Jiang L. Low-dose anticoagulation in Chinese patients with mechanical heart valves. Asian Cardiovasc Thorac Ann 2005 Dec;13(4):341-344. [doi: 10.1177/021849230501300410] [Medline: 16304222]
20. Kearon C, Akl EA, Ornelas J, Blaivas A, Jimenez D, Bounameaux H, et al. Antithrombotic Therapy for VTE Disease: CHEST Guideline and Expert Panel Report. Chest 2016 Feb;149(2):315-352. [doi: 10.1016/j.chest.2015.11.026] [Medline: 26867832]
21. Hart RG, Pearce LA, Aguilar MI. Meta-analysis: antithrombotic therapy to prevent stroke in patients who have nonvalvular atrial fibrillation. Ann Intern Med 2007 Jun 19;146(12):857-867. [doi: 10.7326/0003-4819-146-12-200706190-00007] [Medline: 17577005]
22. Olesen J, Lip G, Lindhardsen J, Lane D, Ahlehoff O, Hansen M, et al. Risks of thromboembolism and bleeding with thromboprophylaxis in patients with atrial fibrillation: A net clinical benefit analysis using a 'real world' nationwide cohort study. Thromb Haemost 2011 Oct;106(4):739-749. [doi: 10.1160/TH11-05-0364] [Medline: 21789337]
23. Wan Y, Heneghan C, Perera R, Roberts N, Hollowell J, Glasziou P, et al. Anticoagulation control and prediction of adverse events in patients with atrial fibrillation: a systematic review. Circ Cardiovasc Qual Outcomes 2008 Nov;1(2):84-91. [doi: 10.1161/CIRCOUTCOMES.108.796185] [Medline: 20031794]
24. Singer DE, Hellkamp AS, Piccini JP, Mahaffey KW, Lokhnygina Y, Pan G, ROCKET AF Investigators. Impact of global geographic region on time in therapeutic range on warfarin anticoagulant therapy: data from the ROCKET AF clinical trial. J Am Heart Assoc 2013 Feb 19;2(1):e000067 [FREE Full text] [doi: 10.1161/JAHA.112.000067] [Medline: 23525418]
25. Prochaska JH, Göbel S, Keller K, Coldewey M, Ullmann A, Lamparter H, et al. Quality of oral anticoagulation with phenprocoumon in regular medical care and its potential for improvement in a telemedicine-based coagulation service--results from the prospective, multi-center, observational cohort study thrombEVAL. BMC Med 2015 Jan 23;13:14 [FREE Full text] [doi: 10.1186/s12916-015-0268-9] [Medline: 25616558]

26. Lee JC, Horner KE, Krummel ML, McDanel DL. Clinical and Financial Outcomes Evaluation of Multimodal Pharmacist Warfarin Management of a Statewide Urban and Rural Population. J Pharm Pract 2018 Apr;31(2):150-156. [doi: 10.1177/0897190017703506] [Medline: 28441899]
27. Cryder BT, Felczak MA, Darkwa A, Patel H, Janociak JD, Rihani R. Transition of stable patients from traditional anticoagulation clinic services to telephonic management. Int J Clin Pharm 2017 Jun;39(3):569-572. [doi: 10.1007/s11096-017-0428-4] [Medline: 28361456]
28. Salvador CH, Ruiz-Sanchez A, González de Mingo MA, Carmona Rodríguez M, Carrasco MP, Sagredo PG, et al. Evaluation of a telemedicine-based service for the follow-up and monitoring of patients treated with oral anticoagulant therapy. IEEE Trans Inf Technol Biomed 2008 Nov;12(6):696-706. [doi: 10.1109/TITB.2008.910750] [Medline: 19000948]
29. Stoudenmire LG, DeRemer CE, Elewa H. Telephone versus office-based management of warfarin: impact on international normalized ratios and outcomes. Int J Hematol 2014 Aug;100(2):119-124. [doi: 10.1007/s12185-014-1619-6] [Medline: 24952031]
30. Blissit KT, Mullenix ML, Brittain KG. Evaluation of Time in Therapeutic Range on Warfarin Therapy Between Face-to-Face and Telephone Follow-Up in a VA Medical Center. Journal of Pharmacy Technology 2014 Oct 12;31(2):78-83. [doi: 10.1177/8755122514552497]
31. Lin S, Kang W, Lin D, Lee J, Wu F, Chen C, et al. Comparison of warfarin therapy clinical outcomes following implementation of an automated mobile phone-based critical laboratory value text alert system. BMC Med Genomics 2014;7 Suppl 1:S13 [FREE Full text] [doi: 10.1186/1755-8794-7-S1-S13] [Medline: 25079609]
32. Staresinic AG, Sorkness CA, Goodman BM, Pigarelli DW. Comparison of outcomes using 2 delivery models of anticoagulation care. Arch Intern Med 2006 May 08;166(9):997-1002. [doi: 10.1001/archinte.166.9.997] [Medline: 16682573]
33. Xia X, Fu J, Wu T, Chen W, Jinhua Z. Comparison of the outcomes of warfarin therapy and economics by online and offline anticoagulation management models: protocol for a randomised controlled trial. BMJ Open 2019 Sep 05;9(9):e031730 [FREE Full text] [doi: 10.1136/bmjopen-2019-031730] [Medline: 31492796]

Abbreviations

FMUUH: Fujian Medical University Union Hospital
INR: international normalized ratio
TTR: time in therapeutic range

Edited by L Buis; submitted 08.08.20; peer-reviewed by M Ghozali, A Manteghinejad, AS Aslam; comments to author 22.10.20; revised version received 13.12.20; accepted 08.01.21; published 02.03.21.

Please cite as:
Cao H, Jiang S, Lv M, Wu T, Chen W, Zhang J
Effectiveness of the Alfalfa App in Warfarin Therapy Management for Patients Undergoing Venous Thrombosis Prevention and Treatment: Cohort Study
JMIR Mhealth Uhealth 2021;9(3):e23332
URL: https://mhealth.jmir.org/2021/3/e23332
doi:10.2196/23332
PMID:33650976

Original Paper

User Engagement and Clinical Impact of the Manage My Pain App in Patients With Chronic Pain: A Real-World, Multi-site Trial

Anuj Bhatia[1,2], MD, FRCPC; Jamal Kara[3,4], BSc, CCRP; Tahir Janmohamed[5], BSc, MBA; Atul Prabhu[1], MBBS, FRCPC; Gerald Lebovic[2,6], BSc, MSc, PhD; Joel Katz[1,7], MA, PhD; Hance Clarke[1,8], MD, PhD, FRCPC

[1]Department of Anesthesia and Pain Medicine, University Health Network, University of Toronto, Toronto, ON, Canada

[2]Institute of Health Policy Management and Evaluation, University of Toronto, Toronto, ON, Canada

[3]Department of Anesthesia and Pain Management, Toronto Western Hospital, University Health Network, Toronto, ON, Canada

[4]Institute of Psychiatry, Psychology, and Neuroscience, King's College London, London, United Kingdom

[5]ManagingLife Inc, Toronto, ON, Canada

[6]Applied Health Research Centre, Unity Health Toronto, Toronto, ON, Canada

[7]Department of Psychology, York University, Toronto, ON, Canada

[8]Transitional Pain Service, Toronto General Hospital, University Health Network, Toronto, ON, Canada

Corresponding Author:
Anuj Bhatia, MD, FRCPC
Department of Anesthesia and Pain Medicine
University Health Network
University of Toronto
McL 2-405
399 Bathurst Street
Toronto, ON, M5T 2S8
Canada
Phone: 1 4166035800 ext 6136
Email: anuj.bhatia@uhn.ca

Abstract

Background: Chronic pain imposes a large burden on individuals and society. A patient-centric digital chronic pain management app called Manage My Pain (MMP) can be used to enhance communication between providers and patients and promote self-management.

Objective: The purpose of this study was to evaluate the real-world engagement of patients in urban and rural settings in Ontario, Canada with the MMP app alongside their standard of care and assess the impact of its usage on clinical outcomes of pain and related mental health.

Methods: A total of 246 participants with chronic pain at a rural and 2 urban pain clinics were recruited into this prospective, open-label, exploratory study that compared the use of MMP, a digital health app for pain that incorporates validated questionnaires and provides patients with summarized reports of their progress in combination with standard care (app group), against data entered on paper-based questionnaires (nonapp group). Participants completed validated questionnaires on anxiety, depression, pain catastrophizing, satisfaction, and daily opioid consumption up to 4.5 months after the initial visit (short-term follow-up) and between 4.5 and 7 months after the initial visit (long-term follow-up). Engagement and clinical outcomes were compared between participants in the two groups.

Results: A total of 73.6% (181/246) of the participants agreed to use the app, with 63.4% (111/175) of them using it for at least one month. Individuals who used the app rated lower anxiety (reduction in Generalized Anxiety Disorder 7-item questionnaire score by 2.10 points, 95% CI –3.96 to –0.24) at short-term follow-up and had a greater reduction in pain catastrophizing (reduction in Pain Catastrophizing Scale score by 5.23 points, 95% CI –9.55 to –0.91) at long-term follow-up relative to patients with pain who did not engage with the MMP app.

Conclusions: The use of MMP by patients with chronic pain is associated with engagement and improvements in self-reported anxiety and pain catastrophizing. Further research is required to understand factors that impact continued engagement and clinical outcomes in patients with chronic pain.

Trial Registration: ClinicalTrials.gov NCT04762329; https://clinicaltrials.gov/ct2/show/NCT04762329

(JMIR Mhealth Uhealth 2021;9(3):e26528) doi:10.2196/26528

KEYWORDS

pain; psychology; patient-oriented research; quality of life; digital health; chronic pain; pain app; virtual care; mHealth; pain management; chronic disease management; remote monitoring; app; engagement; impact; outcome

Introduction

Several large population-based surveys show that over 1 in 5 people live with chronic pain [1-5]. Pain is associated with poor quality of life [6] and is one of the top 3 reasons to seek medical attention in Canada [7,8]. The combined direct and indirect annual costs of chronic pain in North America are estimated to be more than US $650 billion [6,9-11]. Despite these staggering numbers, there are barriers to improving the management and outcomes of chronic pain, including obtaining longitudinal data, assessing response to interventions, and addressing challenges to communication between patients and health care providers (HCPs) [12]. The need to maintain continuity of care for chronic pain patients has also become imperative to avoid treatment disruptions due to public health emergencies, such as COVID-19, limiting in-person visits [13,14].

To bridge this gap, a patient-centric digital health app can be used as a method of remote monitoring to enhance communication between patients and HCPs and promote self-management of patients' symptoms. While a number of pain apps have been created, they evaluate the biopsychosocial components of pain experiences inadequately and lack clinical involvement [12,15-17]. The scientific validation process for these digital pain apps has not focused on development, adoption, engagement, and patient satisfaction [18-21]. None have been scientifically validated for their impact on pain-related clinical outcomes [16,17]. The few digital pain management solutions that have been scientifically validated for positive clinical impact are specific to patients with lower back pain rather than focused on generalized chronic pain, which is prevalent in a multitude of patients with underlying medical conditions [22]. Furthermore, most studies do not address the effectiveness of mobile pain apps based on clinical setting, despite major lifestyle differences of individuals who live in urban and rural areas [23]. Mobile health apps also have difficulty engaging patients, with top-performing health apps having an average 30-day retention rate of only 15% [24].

A novel digital pain management solution, the Manage My Pain (MMP) app (ManagingLife Inc) [25], was used in this study by patients and HCPs to measure and monitor pain, mental health, and medication use. The purpose of this study was to evaluate the real-world engagement of patients in urban and rural settings in Ontario, Canada with the MMP app alongside their standard of care. Engagement was ascertained by evaluating both adoption and retention rates for continuing use of this app over time. Clinical outcomes of pain and related mental health were also measured and compared between patients who engaged with the app versus those who proceeded with the standard of care at their respective institutions.

Methods

Study Sites and Participants

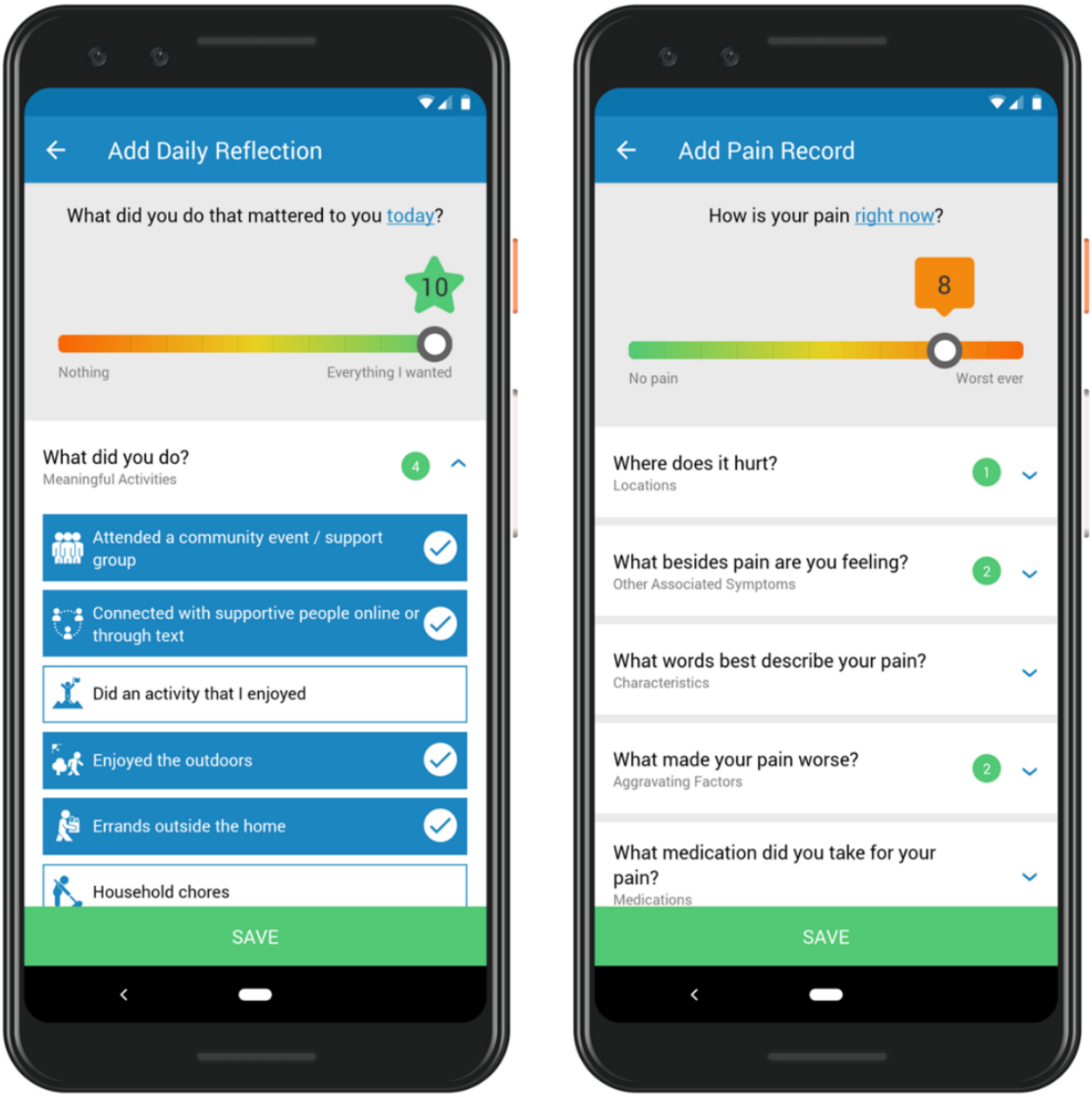
A prospective, open-label, multicenter exploratory study with active and comparator arms was conducted from January 8, 2018, to January 7, 2020, at 3 study sites. Participants were recruited from among new patients with chronic pain conditions who were referred to 2 tertiary academic pain centers in Toronto, Ontario, Canada (Toronto General Hospital [TGH] and Toronto Western Hospital [TWH]) and a rural pain clinic in Ontario (the Iroquois Falls Family Health Team [IFFHT] pain clinic in Iroquois Falls, Ontario). All patients had pain of moderate-to-severe intensity that had persisted for at least three months. All patients, regardless of use of the Manage My Pain app, received the standard of care for the particular clinic, which included interventions such as medication management, psychological therapy, and physiotherapy.

The Manage My Pain Digital Health Solution

Manage My Pain [26], the app used for this study, is a digital health solution that comprises 3 components: (1) an app for patients to track their pain, function, and medication; respond to questionnaires; and view insights on their conditions; (2) reports that summarize the information collected in the app to be used during clinical visits to facilitate communication between patients and clinicians; and (3) a monitoring portal used by clinics to remotely assess patient progress, assign questionnaires, and highlight clinically relevant trends and patterns using advanced analytics [25]. MMP was first launched in 2011 as the first pain management app on the Android platform. In 2015, ManagingLife partnered with the multidisciplinary team to evolve the solution to meet clinical needs and successfully integrate it into the clinical workflow of an outpatient clinic of an academic hospital [27]. Several papers have applied machine learning techniques to analyze the engagement patterns of users within MMP and develop prediction models from its data set of over 50,000 users [28-30].

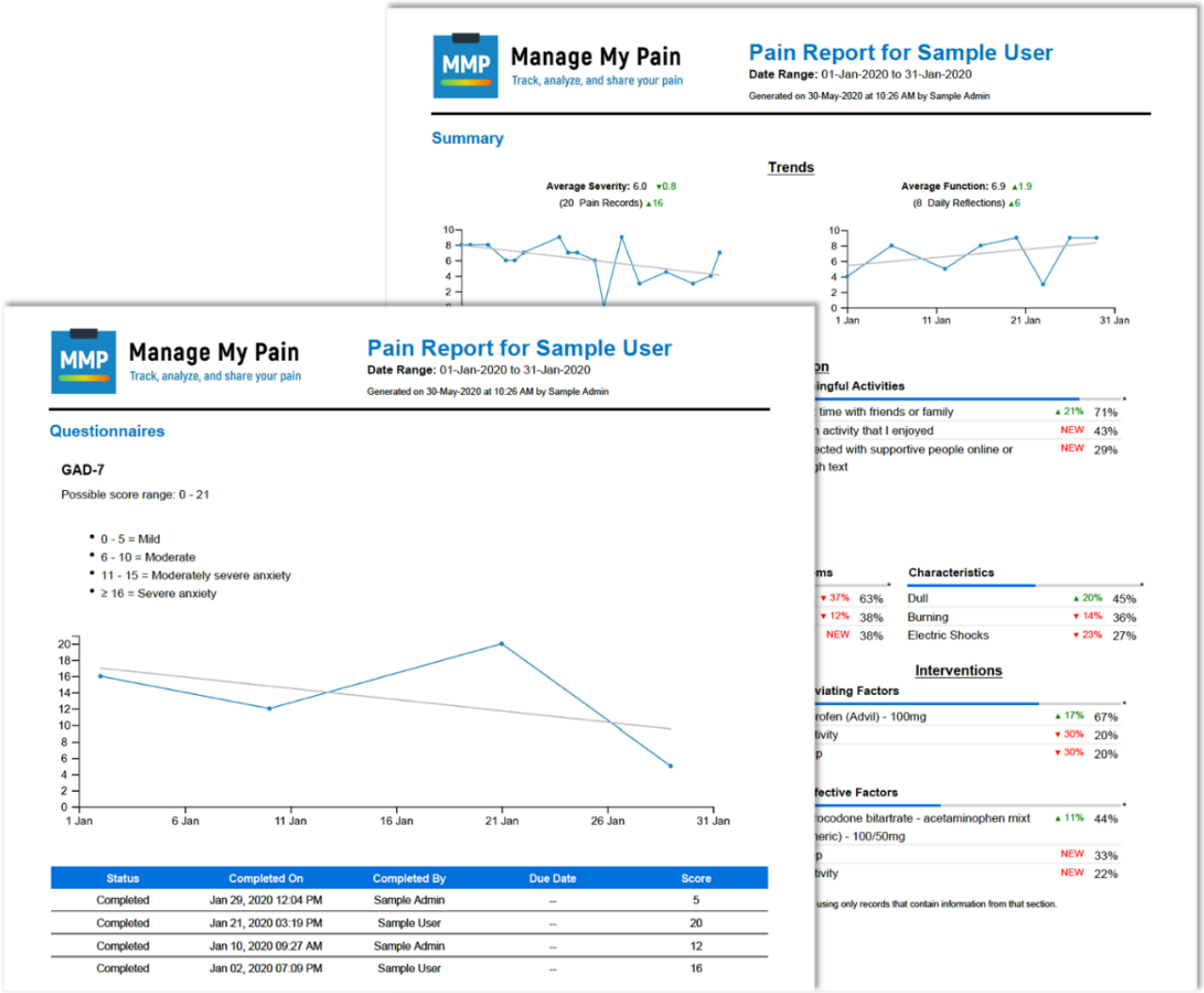
MMP is used by both patients and clinics to measure and monitor pain, function, and medication use. Patients can record their experiences by using the MMP app on their mobile device (compatible with Android and iOS devices) or accessing a web-based platform. When prompted by an in-app push notification triggered at 8 PM daily, patients record daily reflections in the app, where they indicate the meaningful activities they were able to accomplish. The daily reflection concept is based on acceptance and commitment therapy principles that have demonstrated an ability to improve clinical outcomes relevant to pain management [27]. Patients also record their pain episodes, including descriptions such as severity, locations, associated symptoms, characteristics, duration, environment, and aggravating or alleviating factors (Figure 1).

Figure 1. Screenshots of the screens used by Manage My Pain to collect patient-reported outcomes.



Each section of the app can be customized by the patient so that it is representative of their situation. MMP is designed to capture both daily reflections and pain episodes in less than 60 seconds [25]. As patients enter information into the app, charts and graphs are presented to the patient to highlight patterns and trends that increase self-awareness of their condition and provide insight into triggers and interventions. With consent, this information can also be viewed by their clinical team through MMP's remote monitoring portal. For clinics, pain and function trends are summarized across a predefined time period and can be viewed digitally or output into a clinician-friendly concise report. These self-reported outcomes are used to improve communication with a patient during a clinical visit and assess progress more objectively between clinical visits. Moreover, MMP allows clinics to assign validated questionnaires on pain and related domains for patients to complete at home in advance of their clinical visit. Responses to these questionnaires, along with their corresponding scores and interpretations, are also available through the MMP portal and can be summarized in the clinician-facing report (Figure 2). Emails are sent to the patients by MMP at predefined intervals to encourage engagement with the app and prompt patients to complete the questionnaires by the specified due date.

Figure 2. Sample clinician-facing report produced by Manage My Pain that contains scored responses to validated measures as well as a summary of the daily tracking.



Recruitment

As part of the registration process, patients were shown a consent screen asking for permission to share the information from MMP with their clinical team for clinical and research purposes, and they signed an informed consent form. Participants had to explicitly agree by entering a unique randomized 10-digit ID provided to them by the research coordinator, which allowed each clinic to match the participant with their clinical profile and ensure appropriate deidentification. After registration, participants had to activate their account by clicking on a link sent to them by email. Participants were instructed to download MMP through either the Apple App Store (for iPhone users) or the Google Play Store (for Android users) upon successful activation. For participants that did not have either device, a web link associated with their clinic was bookmarked for them to use MMP through their internet browser for easy access.

At first access, the app provided a brief tutorial on how to set up the user's profile, which included entering relevant medications and pain conditions. It then described how to complete a sample entry by using a touch slider to enter a numerical rating score as an integer from 0 to 10. Once the user's pain level was indicated, additional questions were prompted, including pain location, other associated symptoms, pain characteristics, aggravating and alleviating factors, and the duration and environment of the pain episode. While all questions included a drop-down menu including prepopulated answers with associated infographics, participants were also given the option to add customized values. A Notes section where participants could enter free text was also provided. Upon completion, patients could save their entry for the patient and health care team to track and observe trends. Technical support was also offered to the participants in the event that they required troubleshooting for using the app, either by the designated research team or the ManagingLife technical support team. Email notifications for data entry on validated questionnaires were sent by the app 7 days prior to the due date set by the research coordinator, which was intended to coincide with the date of the clinical visit. This feature allowed HCPs to track the progress of patients who could not be seen through in-person visits.

Study Procedures

Participants at the 3 sites were selected using a homogeneous purposive sampling method [31] and given an option to use the

app. During the first clinical visit, patients who consented to the study for their data to be collected were offered a choice to either register an account with the app [25] to share their data with the research and clinical teams through the app's monitoring portal or provide their data on paper-based questionnaires during pain clinic visits. Participants who continued to record their pain experiences and enter data in the app after 30 days of first registering were considered to be in the app group. Participants who declined to use the app or those who registered but had no records in the app after 30 days were considered to be in the nonapp group. A cutoff of 30 days was used based on its acceptance in the mobile app industry as a benchmarked metric of retention [24] and on its determination by the clinical team involved in the study as a meaningful duration of information that could inform clinical decision making.

Participants were asked to complete questionnaires on the following pain-related outcomes: anxiety, depression, pain catastrophizing, pain disability, patient global impression of change, and daily opioid consumption. Daily opioid consumption was measured in oral morphine equivalents in milligrams upon entry into the study during the initial visit and as a part of the first follow-up clinical visit within 4.5 months of the initial visit (short-term follow-up) and during the second follow-up clinical visit between 4.5 and 7 months after the initial visit (long-term follow-up). Given that our objective was to validate the impact of the app in a real-world clinical environment, the date ranges of follow-up visits were more broadly defined to align with the date of the actual clinical visit. Clinicians at each participating site were encouraged to use the clinical reports, either digitally through the portal or printed, during their clinical visits with the patients.

Participants who agreed to use the app but did not have in-person clinical appointments within these time frames were still remotely prompted to complete the questionnaires through the app portal by research staff. Patients in the nonapp group provided their data only by completing paper-based questionnaires during clinic visits or via a phone interview if no in-person clinical visit was scheduled during study-related follow-ups.

Validated Measures Used in the Study

The feasibility and successful adoption of the digital health solution was evaluated through clinical outcomes and patient engagement. During the initial visit, participants completed baseline questionnaires that were standardized across sites as well as questionnaires that were considered the standard of care at each individual clinical practice.

As mood disorders are often prevalent in individuals with chronic pain and have been known to affect and intensify pain perception [32,33], anxiety and depression levels were recorded using the Generalized Anxiety Disorder 7-item questionnaire (GAD-7) and the Patient Health Questionnaire 9-item scale (PHQ-9) for depression, respectively. These questionnaires have repeatedly demonstrated excellent test-retest reliability [34], criterion and construct validity [35], and high levels of specificity and sensitivity in the assessment of anxiety and depression in patients with chronic pain [34-38]. In addition, the score on the Pain Disability Index (PDI), a 7-item instrument used to evaluate the degree of pain-related disability, can be inversely correlated with overall function [39]. The Pain Catastrophizing Scale (PCS) was also administered to participants to measure the degree of maladaptive cognitive distortions known as catastrophizing, which increase negative emotional schema throughout the anticipation and experience of painful stimulation [40,41]. Additionally, the Patient Global Impression of Change (PGIC) questionnaire, a 7-level ordinal measure, was administered to participants at both short-term and long-term follow-up time frames to assess the degree of improvement or worsening of a patient's clinical condition. The PGIC is a single-item validated questionnaire that asks the user about their perceived improvement and is significantly correlated with changes in pain intensity, efficacy of treatment, and interference of pain in daily activities [42].

Finally, participants' opioid consumption was measured over time using oral morphine equivalents (OME) through conversion ratios outlined by the Canadian Guideline for Safe and Effective Use of Opioids for Chronic Non-Cancer Pain [43]. Given that the core functionality of Manage My Pain is to record pain intensity scores during each engagement with the app, it was not selected as a measure for evaluation, as any comparisons against patients recording this information using point-in-time questionnaires would be misleading. Specifically, other measures were selected to assess the mental and physical well-being of patients independent of their severity and intensity.

Ethics Approval

Institutional ethics board approvals were obtained from each study site by the University Health Network Research Ethics Board for the academic sites (TGH and TWH) and Veritas Institutional Review Board for the rural site (IFFHT). The approval process involved confirmation that MMP has the administrative and technical safeguards in place to ensure compliance with privacy legislation.

Statistical Analyses

Continuous data were summarized using mean and standard deviation or median and interquartile ranges, and categorical data were summarized using frequency and percentages. Univariate tests for continuous data were conducted using 2-sample *t* tests or Wilcoxon rank sum tests as appropriate based on the distribution of the data. Chi-square tests or Fisher exact tests were used for categorical data. A random-effects model was used for all outcomes to account for correlations arising from repeated measures within the same individual.

The main exposure of interest was whether someone used the app for at least 30 days, adjusted for time, age, gender, and study site. Each of the 6 outcomes (daily OME, GAD-7 score, PHQ-9 score, PDI score, PCS score, and PGIC score) was also modeled to examine the association of the intervention (use of the app for at least 30 days) with the outcome after controlling for other relevant variables. The duration of usage (time) among the participants who entered data into the app was recorded as the difference between the most recent entry and the date of registration. An interaction term between time (short term or long term) and group (intervention or control) was examined

to determine whether the intervention was associated with differences between groups over time. A likelihood ratio test was used to assess the statistical significance of the interaction term, and the term was included in the model if it remained statistically significant at the .05 significance level. For clinical utility, the primary analysis categorized time as baseline, short term, and long term. A sensitivity analysis was conducted using time as a continuous covariate of interest. The baseline value was adjusted for by including it in the outcome vector [44]. Model checking for continuous outcomes was performed using analysis of residuals. Bootstrapped 95% confidence intervals and *P* values were provided where the model residuals violated the normality assumption.

Results

Data at Baseline and Engagement With the App

The average age of participants in the study was 56.67 (SD 13.12) years, with 60.2% (148/246) of participants being female. A total of 246 participants were enrolled across the 3 sites (154 participants at the 2 urban sites and 92 participants at the rural site), out of which 181 (73.6%) accepted the use of the app in their clinical care and the remaining 65 (26.4%) continued with paper-based data entry at their respective clinics. Of the 181 participants who agreed to use the app, 175 (96.7%) participants registered and provided consent to share their data. Of the 175 participants who registered, 111 (63.4%) participants used the app for at least 30 days and therefore were considered part of the intervention (app) group (Figure 3). Data from 70 participants who initially accepted the use of the app but either did not use the app or used it for less than 30 days were combined with data from the 65 participants who had declined to use the app at the start of the study, and these 135 participants were considered to be in the nonapp group for analysis (Figure 3). There were no differences between the app and nonapp groups with respect to demographics, duration of pain, or the validated measures for mood and physical disability (Table 1).

Figure 3. Flow diagram of group allocation based on participants' engagement with the app.

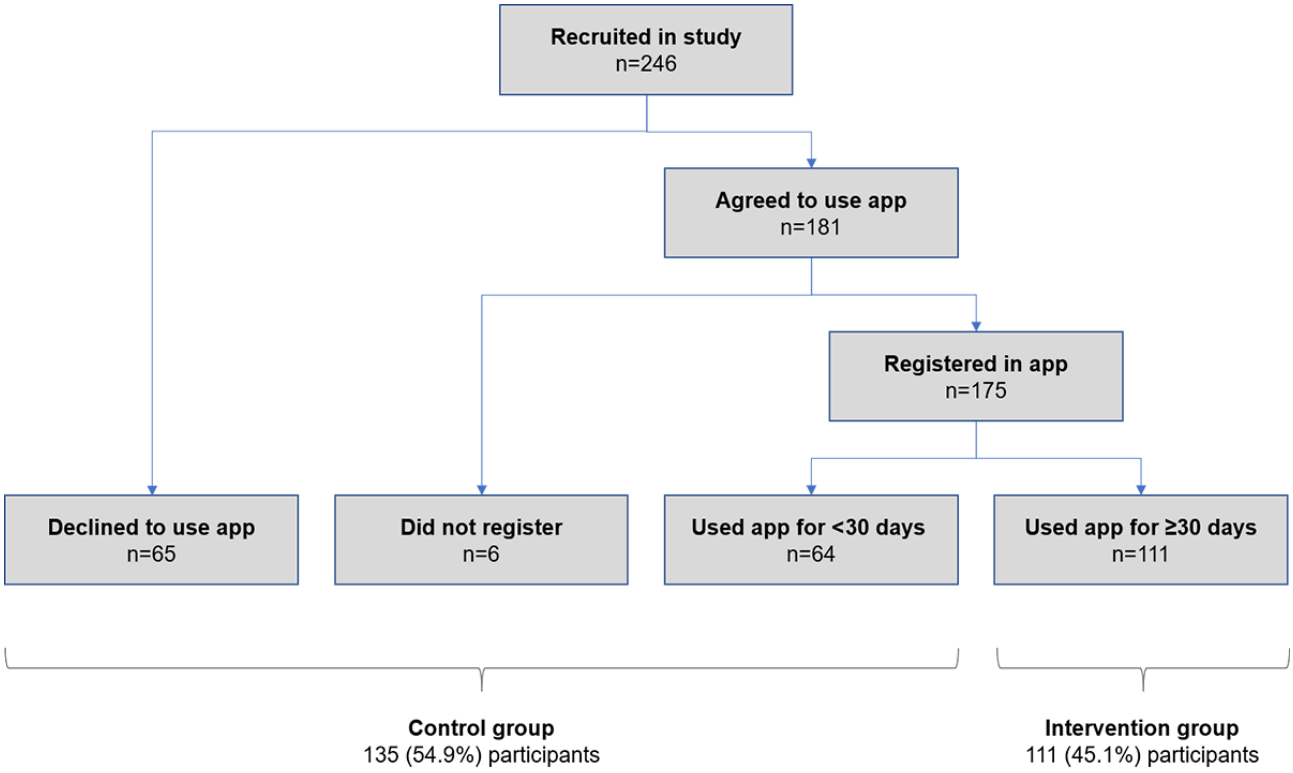


Table 1. Baseline demographics characteristics of the study population.

Characteristic	Nonapp group[a] (n=135)	App group (n=111)	*P* value
Age (years), mean (SD)	57.17 (15.07)	56.05 (10.31)	.51
Male sex, n (%)	59 (44.0)	39 (35.1)	.20
Employment status, n (%)			.69
Never worked	1 (0.8)	0 (0.0)	
Not working	58 (45.3)	51 (46.4)	
Working (part- or full-time)	27 (21.1)	27 (24.5)	
Retired	42 (32.8)	32 (29.1)	
Duration of pain, n (%)			.16
<12 months	7 (5.3)	13 (11.9)	
12-24 months	13 (9.8)	12 (11.0)	
>24 months	112 (84.8)	84 (77.1)	
Etiology of pain, n (%)			.005
Accident	35 (26.9)	35 (31.8)	
As a result of illness	27 (20.8)	11 (10.0)	
Following surgery	13 (10.0)	8 (7.3)	
No known reason	43 (33.1)	54 (49.1)	
Other	12 (9.2)	2 (1.8)	
GAD-7[b] score, mean (SD)	8.51 (6.21)	8.21 (6.35)	.71
PHQ-9[c] score, mean (SD)	11.26 (6.83)	11.52 (6.66)	.77
PDI[d] score, mean (SD)	40.30 (16.47)	41.76 (15.11)	.49
PCS[e] score, mean (SD)	23.84 (13.19)	24.93 (14.39)	.55
OME[f] (mg/24 hr), median (IQR)	0 (0-30)	0 (0-27)	.42

[a]The nonapp group included data from the 65 participants who had declined to use the app at the start of the study and the 70 participants who initially accepted use of the app but either did not use the app or used it for less than 30 days.

[b]GAD-7: Generalized Anxiety Disorder 7-item scale.

[c]PHQ-9: Patient Health Questionnaire 9-item scale.

[d]PDI: Pain Disability Index.

[e]PCS: Pain Catastrophizing Scale.

[f]OME: oral morphine equivalents.

Of those who used the app for at least 30 days, the mean number of records entered was 113.4 (SD 129.7). The mean duration of usage from the date of registration to the date of the last record entered was 164.2 (SD 88.4) days.

Clinical Outcomes

Quantitative analysis of the clinical outcomes was performed at 2 time points, short term and long term, following enrollment into the study. Of the 135 patients in the nonapp group, 36 (26.7%) provided their data at the short-term follow-up and 31 (23.0%) provided their data at the long-term follow-up. A total of 90 of the 111 participants in the app group provided their data through the app for the short-term follow-up, and 69 provided their data for both short- and long-term follow-ups (Table 2). The primary reason for the large number of patients whose data were not available at the short-term follow-up is that many were deemed not to qualify for treatments offered at the clinics shortly after consenting to participate in the study and therefore were discharged from the clinic. The patients were discharged because these clinics accepted only patients with pain whose condition was amenable to the interventions offered at the clinic (eg, high-dose intravenous ketamine infusions, neuromodulation implants). For each of the measures collected from participants at the short-term and long-term follow-ups, less than 5% of data were missing for each time point, and no imputation technique was used.

The number of elapsed days from baseline for both the short-term and long-term follow-ups was not significantly different across the intervention and control groups (Table 3).

Unadjusted analyses did not find any significant differences between the intervention and control groups over time (Table 4 and Figure 4).

Table 2. Number of participants with questionnaire responses at the short-term follow-up (prior to 4.5 months from baseline) and the long-term follow-up (between 4.5 and 7 months from baseline).

Group	Participants, n	Baseline, n (%)	Short-term, n (%)	Long-term, n (%)
Nonapp	135	130 (96.3)	36 (27.7)	31 (23.8)
Declined to use app	65	64 (98.5)	22 (34.4)	19 (29.7)
Did not register	6	3 (50.0)	2 (66.7)	2 (66.7)
Used app for <30 days	64	63 (98.4)	12 (19.0)	10 (15.9)
App (used app for ≥30 days)	111	111 (100.0)	90 (81.1)	69 (62.2)
Total	246	241 (98.0)	126 (52.3)	100 (41.5)

Table 3. Days from baseline for both short-term and long-term follow-up time periods. *P* values used a Wilcoxon test and 95% CIs were bootstrapped.

Time	App group, median (IQR) (n=111)	Nonapp group, median (IQR) (n=135)	Difference (95% CI)	*P* value
Short-term	92 (80-100)	91 (78-104)	1 (–7 to 8)	.80
Long-term	183 (162-197)	188.5 (168-194)	–5.5 (–18 to 2)	.83

Figure 4. Mean values of validated measures in both app (labelled "Intervention – Yes") and nonapp (labelled "Intervention – No") groups at short-term (early) and long-term (late) follow-ups. Error bars indicate the standard error. Means have been centered along the overall mean. GAD-7: Generalized Anxiety Disorder 7-item scale; OME: oral morphine equivalence; PCS: Pain Catastrophizing Scale; PDI: Pain Disability Index; PGIC: Participant Global Impression of Change; PHQ-9: Patient Health Questionnaire 9-item scale.

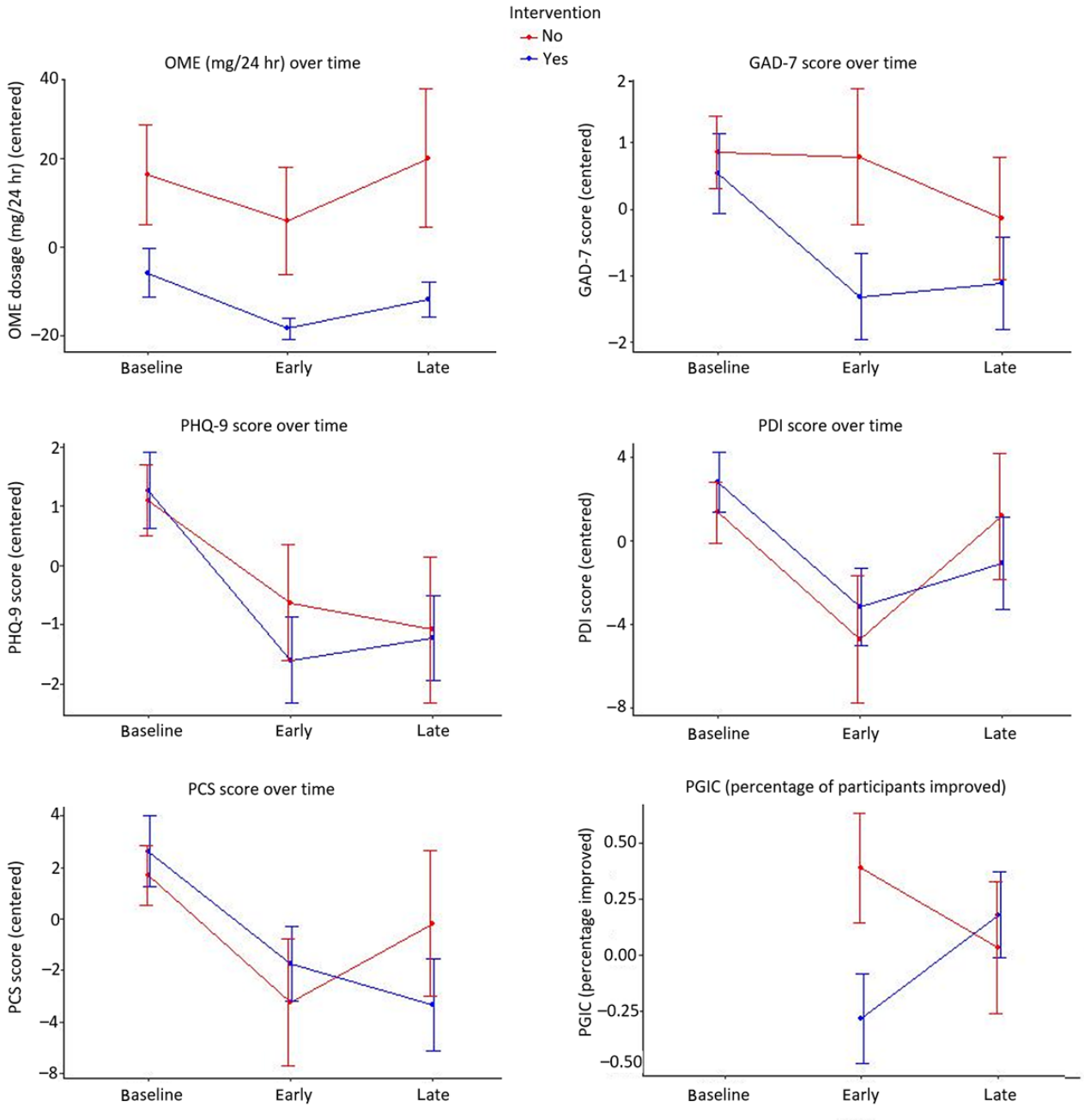


Table 4. Unadjusted outcomes stratified by time and group. The differences between the intervention and control groups are represented as absolute numbers with their 95% CIs, and 95% CI for OME used bootstrapped values.

Time	App group	Nonapp group	Difference (95% CI)	*P* value
OME[a] (mg/24 hr), median (IQR)				
Baseline	0 (0-27)	0 (0-30)	0 (–14 to 0)	.39
Short-term	0 (0-15)	0 (0-27)	0 (–30 to 0)	.18
Long-term	0 (0-23)	2 (0-73)	–2 (–71 to 0)	.16
GAD-7[b] score, mean (SD)				
Baseline	8.21 (6.35)	8.53 (6.23)	–0.32 (–1.95 to 1.30)	.70
Short-term	6.34 (6.17)	8.46 (6.13)	–2.12 (–4.58 to 0.35)	.09
Long-term	6.54 (5.86)	7.53 (5.14)	–0.99 (–3.34 to 1.36)	.43
PHQ-9[c] score, mean (SD)				
Baseline	11.52 (6.66)	11.35 (6.78)	0.17 (–1.57 to 1.91)	.85
Short-term	8.66 (6.93)	9.63 (5.92)	–0.97 (–3.46 to 1.53)	.47
Long-term	9.03 (5.88)	9.17 (6.84)	–0.14 (–3.02 to 2.75)	.92
PDI[d] score, mean (SD)				
Baseline	41.76 (15.11)	40.32 (16.53)	1.44 (–2.67 to 5.55)	.49
Short-term	35.78 (17.57)	34.23 (18.37)	1.55 (–5.74 to 8.83)	.67
Long-term	37.90 (18.32)	40.14 (16.74)	–2.24 (–10.01 to 5.52)	.58
PCS[e] score, mean (SD)				
Baseline	24.93 (14.39)	24.00 (13.13)	0.93 (–2.65 to 4.51)	.61
Short-term	20.58 (13.77)	19.06 (14.84)	1.52 (–4.32 to 7.36)	.59
Long-term	18.97 (14.94)	22.13 (15.80)	–3.16 (–10.00 to 3.67)	.35
PGIC[f] score[g], n/N (%)				
Baseline	N/A[h]	N/A	N/A	N/A
Short-term (improved)	32/89 (36.0)	15/32 (46.9)	–10.9% (–33.0% to 11.1%)	.38
Long-term (improved)	24/58 (41.4)	11/28 (39.3)	2.1% (–22.1% to 26.3%)	>.99

[a]OME: oral morphine equivalents.

[b]GAD-7: Generalized Anxiety Disorder 7-item scale.

[c]PHQ-9: Patient Health Questionnaire 9-item scale.

[d]PDI: Pain Disability Index.

[e]PCS: Pain Catastrophizing Scale.

[f]PGIC: Patient Global Impression of Change.

[g]PGIC score represents participants who improved.

[h]N/A: not applicable.

Adjusted Analysis for the Entire Study Cohort

A significant decline in daily OME in milligrams was observed in both the short-term (decrease of 8.31 mg, 95% CI –16.62 to –0.97) and long-term (decrease of 12.59 points, 95% CI –21.16 to 4.27) time periods when compared with baseline. Depression (PHQ-9) (lower by 2.29 points, 95% CI –3.23 to –1.34 in the short-term follow-up; lower by 2.52 points, 95% CI –3.56 to –1.47 in the long-term follow-up) and disability (PDI) scores (lower by 5.20 points, 95% CI –7.60 to –2.81 in the short-term follow-up; lower by 3.52 points, 95% CI –6.20 to –0.80 in the long-term follow-up) decreased significantly for all participants (Multimedia Appendix 1). Pain Catastrophizing Scale scores decreased at the short-term follow-up, with the scores lowered by 3.53 points (95% CI –6.88 to –0.17) but returned to baseline at the long-ter m follow-up. Older participants reported lower opioid use over time, with a decrease of 0.98 mg of OME per year of increasing age (95% CI –1.80 to –0.08). Increasing age was also associated with lower GAD-7, PHQ-9, PDI, and PCS scores (Multimedia Appendix 1). Male sex was associated with a higher disability score throughout the study (Multimedia Appendix 1).

Participants in the app group had lower anxiety (GAD-7) scores at the short-term follow-up (decrease of 2.10 points, 95% CI –3.96 to –0.24) and lower Pain Catastrophizing Scale scores at the long-term follow-up (decrease of 5.23 points, 95% CI –9.55 to –0.91) (Multimedia Appendix 1). For the reduction in the anxiety and pain catastrophizing scores, there was a significant intervention-by-time interaction, indicating that the decrease in these scores was higher in the group that used the app. There was also a change over time for the daily OME (lower by 12.59 mg, 95% CI –21.16 to –4.27), PHQ-9 score (decrease of 2.52 points, 95% CI –3.56 to –1.47), and PDI score (decrease of 3.52 points, 95% CI –6.20 to –0.80), but there was no time-by-intervention interaction, indicating that the change in outcome over time was not different between those who used the app and those who did not.

Discussion

Summary of the Main Results of the Study

This is the first multisite study at rural and urban pain clinics of a digital pain management solution that compared outcomes in patients with chronic pain who chose to use the app in addition to standard care versus those who received only standard care. A total of 73.6% (181/246) of the participants in the study chose to enroll for the app, 45.1% (111/246) continued to use it beyond one month following enrollment, and 28.0% (69/246) were still using it for 4.5 to 7 months. There was evidence of a decrease in anxiety and pain catastrophizing in participants who used the app versus those who did not use the app.

Acceptance of and Engagement With the App

Digital health applications can play a significant role in enhancing the connectivity between patients and their HCPs. Though some studies report men and younger age groups as more likely to engage with this kind of technology [45], our study and others did not find differences based on age and sex [45,46]. It is possible that the chronicity of pain in participants in our study and the lack of effective therapies made patients in our study interested in exploring the potential for help from the app offered in our study. The rates of initial engagement with the app reported in our study—initially 73.6% (181/246), with a gradual decrease to 28.0% (69/246) at the long-term follow-up—are consistent with those reported in literature by our group [47] and others [45,48,49] and appear to be better than the rates for other apps [24]. We measured ongoing engagement with the digital app, unlike other studies that evaluate merely the intent of patients to engage with digital health solutions [46]. Continuing engagement with digital health solutions is important, and international health organizations have also emphasized the importance of developing evidence for the integration of digital health solutions in routine medical care [27,28] because of their potential to empower and enable patients. A follow-up study will focus on the engagement patterns and their contributing factors within this study along with their potential correlations with the clinical outcomes seen.

Association of Using the App With Pain-Related Clinical Outcomes

Our study found that use of the app was associated with a reduction in pain-related anxiety and pain catastrophizing scores. These reductions have clinical significance, given that the minimum clinically important difference in values for the GAD-7 and PCS (–4 for GAD-7 [50] and 38% for PCS [51]) is within the 95% confidence interval of the outcome reduction. Pain-related anxiety and catastrophizing can have significant adverse effects on patients, with an increase in both health care use [52] and the probability of misuse of prescription opioids [53].

The ability of patients to track and reflect on their pain and its relationship to activities and medications in our study may have resulted in an attenuation of the psychological correlates of chronic pain. Self-monitoring of symptoms is an important component of most pain self-management programs [54]. There is a growing body of evidence that self-monitoring using eHealth tools is associated with positive health outcomes [55]. In particular, the daily reflection concept used by Manage My Pain is a form of self-monitoring based on acceptance and commitment therapy principles, which is an empirically supported treatment for individuals living with chronic pain [56]. It emphasizes engagement in meaningful activities based on personal values as a cornerstone of treatment [57]. The use of the app's diary of patients' lived experiences when interacting with HCPs through reports or the remote monitoring portal possibly empowered patients to address their negative emotions. This empowerment has been known to be associated with an analgesic benefit over time in patients with chronic pain [58,59]. Studies on other chronic health conditions have also reported similar results [60]. While the differences in the other clinical domains between the intervention and control groups were not significant, it is possible that these differences would be significant if the sample size were larger. Further, we did not include pain intensity scores in our study because this instrument has been shown to lack the ability to demonstrate functional benefits of analgesic interventions [61].

This study suggests that engagement by patients with an app-based digital pain solution that incorporates validated questionnaires may be associated with improvement in clinical outcomes. A future study will present the results of a qualitative analysis that assessed both the patients' and clinical team's perspective on the app's utility. Additionally, further research is required to understand factors that impact initial acceptance and continuing engagement with digital apps in patients with pain, including user comfort, understanding of technology, accessibility for patients and connectivity with HCPs, and feasibility of implementation in established health care systems [62].

Limitations of This Study

This study of the outcomes of the use of an app-based digital solution for generalized chronic pain has some limitations. Participants in this study were not allocated to study groups by randomization. This could have introduced a bias because patients comfortable with digital technology were more likely to opt to use the app. An additional bias could have been

introduced in that patients who chose to use the app would have seen clinical improvements regardless of app usage. The high drop-off rate from baseline to the short-term follow-up with the control group may have also introduced a bias in the results. However, the fact that that a large number of patients in this control group were deemed not to qualify for treatments offered at the clinics and were therefore discharged implies that they would not have benefited relative to those at least receiving the standard of care. This may have also contributed to a higher dropout rate for using the app. All measures were based on participant self-reports, which were not verified by objective means (eg, clinical interview to assess anxiety and depressive disorders, verification of opioid use by pharmacy records). This may have resulted in biased estimates of results that differed by treatment group, confounding the present findings.

Conclusions

This study of a novel digital pain management solution that incorporated validated measures for domains of pain in patients at urban and rural clinics found that 28.0% (69/246) of all patients continued to use the app on a long-term basis. Patients that engaged with the digital health solution had less anxiety and lower pain catastrophizing scores as measured by validated tools. Digital pain management applications and other health-related clinical applications deserve significant attention in the years ahead, given the push toward mobile health tools and telemedicine.

Acknowledgments

The authors would like to acknowledge the support from the Department of Anesthesia and Pain Medicine at the University Health Network, Toronto, Ontario, Canada. The authors would also like to thank Dr Stephen Chiang and Dr Auri Bruno-Petrina of the Iroquois Falls Family Health Team.

Conflicts of Interest

TJ is the founder and CEO of ManagingLife Inc, the owner of Manage My Pain. The other authors declare no competing interests or conflicts of interest.

Multimedia Appendix 1
Results of the adjusted analysis.
[DOCX File , 18 KB - mhealth_v9i3e26528_app1.docx]

References

1. Schopflocher D, Taenzer P, Jovey R. The prevalence of chronic pain in Canada. Pain Res Manag 2011;16(6):445-450 [FREE Full text] [doi: 10.1155/2011/876306] [Medline: 22184555]
2. Reitsma M, Tranmer J, Buchanan D, Vandenkerkhof E. The prevalence of chronic pain and pain-related interference in the Canadian population from 1994 to 2008. Chronic Dis Inj Can 2011 Sep;31(4):157-164. [doi: 10.24095/hpcdp.31.4.04]
3. Steingrímsdóttir Ó, Landmark T, Macfarlane GJ, Nielsen CS. Defining chronic pain in epidemiological studies: a systematic review and meta-analysis. Pain 2017 Nov;158(11):2092-2107. [doi: 10.1097/j.pain.0000000000001009] [Medline: 28767506]
4. Jackson T, Thomas S, Stabile V, Shotwell M, Han X, McQueen K. A Systematic Review and Meta-Analysis of the Global Burden of Chronic Pain Without Clear Etiology in Low- and Middle-Income Countries: Trends in Heterogeneous Data and a Proposal for New Assessment Methods. Anesth Analg 2016 Sep;123(3):739-748. [doi: 10.1213/ANE.0000000000001389] [Medline: 27537761]
5. Dahlhamer J, Lucas J, Zelaya C, Nahin R, Mackey S, DeBar L, et al. Prevalence of Chronic Pain and High-Impact Chronic Pain Among Adults - United States, 2016. Morb Mortal Weekly Rep 2018 Sep 14;67(36):1001-1006 [FREE Full text] [doi: 10.15585/mmwr.mm6736a2] [Medline: 30212442]
6. Hogan M, Taddio A, Katz J, Shah V, Krahn M. Incremental health care costs for chronic pain in Ontario, Canada: a population-based matched cohort study of adolescents and adults using administrative data. Pain 2016 Aug;157(8):1626-1633. [doi: 10.1097/j.pain.0000000000000561] [Medline: 26989805]
7. St Sauver JL, Warner DO, Yawn BP, Jacobson DJ, McGree ME, Pankratz JJ, et al. Why patients visit their doctors: assessing the most prevalent conditions in a defined American population. Mayo Clin Proc 2013 Jan;88(1):56-67 [FREE Full text] [doi: 10.1016/j.mayocp.2012.08.020] [Medline: 23274019]
8. Finley CR, Chan DS, Garrison S, Korownyk C, Kolber MR, Campbell S, et al. What are the most common conditions in primary care? Systematic review. Can Fam Physician 2018 Nov;64(11):832-840 [FREE Full text] [Medline: 30429181]
9. Wilson MG, Lavis JN, Ellen ME. Supporting chronic pain management across provincial and territorial health systems in Canada: Findings from two stakeholder dialogues. Pain Res Manag 2015;20(5):269-279 [FREE Full text] [doi: 10.1155/2015/918976] [Medline: 26291124]
10. Lynch ME. The need for a Canadian pain strategy. Pain Res Manag 2011;16(2):77-80 [FREE Full text] [doi: 10.1155/2011/654651] [Medline: 21499581]

11. Gaskin DJ, Richard P. The economic costs of pain in the United States. J Pain 2012 Aug;13(8):715-724. [doi: 10.1016/j.jpain.2012.03.009] [Medline: 22607834]
12. Zhao P, Yoo I, Lancey R, Varghese E. Mobile applications for pain management: an app analysis for clinical usage. BMC Med Inform Decis Mak 2019 May 30;19(1):106 [FREE Full text] [doi: 10.1186/s12911-019-0827-7] [Medline: 31146739]
13. Kapoor A, Guha S, Kanti Das M, Goswami KC, Yadav R. Digital healthcare: The only solution for better healthcare during COVID-19 pandemic? Indian Heart J 2020;72(2):61-64 [FREE Full text] [doi: 10.1016/j.ihj.2020.04.001] [Medline: 32534691]
14. 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 May;161(5):889-893 [FREE Full text] [doi: 10.1097/j.pain.0000000000001885] [Medline: 32251203]
15. Portelli P, Eldred C. A quality review of smartphone applications for the management of pain. Br J Pain 2016 Aug;10(3):135-140 [FREE Full text] [doi: 10.1177/2049463716638700] [Medline: 27583140]
16. Alexander JC, Joshi GP. Smartphone applications for chronic pain management: a critical appraisal. J Pain Res 2016;9:731-734 [FREE Full text] [doi: 10.2147/JPR.S119966] [Medline: 27713649]
17. Salazar A, de Sola H, Failde I, Moral-Munoz JA. Measuring the Quality of Mobile Apps for the Management of Pain: Systematic Search and Evaluation Using the Mobile App Rating Scale. JMIR Mhealth Uhealth 2018 Oct 25;6(10):e10718 [FREE Full text] [doi: 10.2196/10718] [Medline: 30361196]
18. 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 Jan 25;19(1):24 [FREE Full text] [doi: 10.1186/s12911-019-0741-z] [Medline: 30683106]
19. Jamison RN, Mei A, Ross EL. Longitudinal trial of a smartphone pain application for chronic pain patients: Predictors of compliance and satisfaction. J Telemed Telecare 2016 Nov 10;24(2):93-100. [doi: 10.1177/1357633x16679049]
20. Ross EL, Jamison RN, Nicholls L, Perry BM, Nolen KD. Clinical Integration of a Smartphone App for Patients With Chronic Pain: Retrospective Analysis of Predictors of Benefits and Patient Engagement Between Clinic Visits. J Med Internet Res 2020 Apr 16;22(4):e16939 [FREE Full text] [doi: 10.2196/16939] [Medline: 32297871]
21. Jamison RN, Jurcik DC, Edwards RR, Huang C, Ross EL. A Pilot Comparison of a Smartphone App With or Without 2-Way Messaging Among Chronic Pain Patients: Who Benefits From a Pain App? Clin J Pain 2017 Aug;33(8):676-686 [FREE Full text] [doi: 10.1097/AJP.0000000000000455] [Medline: 27898460]
22. Priebe JA, Haas KK, Moreno Sanchez LF, Schoefmann K, Utpadel-Fischler DA, Stockert P, et al. Digital Treatment of Back Pain versus Standard of Care: The Cluster-Randomized Controlled Trial, Rise-uP. J Pain Res 2020 Jul;13:1823-1838. [doi: 10.2147/jpr.s260761]
23. Roberts A, Philip L, Currie M, Mort A. Striking a balance between in-person care and the use of eHealth to support the older rural population with chronic pain. Int J Qual Stud Health Well-being 2015;10:27536 [FREE Full text] [doi: 10.3402/qhw.v10.27536] [Medline: 26336996]
24. Adjust Global App Trends 2019: Global Benchmark Reports. Adjust. 2019. URL: https://www.adjust.com/resources/ebooks/adjust-global-app-trends-report-2019/ [accessed 2021-02-19]
25. ManagingLife. ManagingLife Inc. URL: https://managinglife.com/ [accessed 2020-01-15]
26. Manage My Pain: Track, analyze, and monitor your pain. ManagingLife Inc. URL: https://managemypainapp.com/ [accessed 2021-02-20]
27. Weinrib A, Azam MA, Latman VV, Janmohamed T, Clarke H, Katz J. Manage my pain: A patient-driven mobile platform to prevent and manage chronic postsurgical pain. In: Chronic Illness and Long-Term Care: Breakthroughs in Research and Practice. Hershey, PA: IGI Global; 2018:642-667.
28. Rahman QA, Janmohamed T, Pirbaglou M, Clarke H, Ritvo P, Heffernan JM, et al. Defining and Predicting Pain Volatility in Users of the Manage My Pain App: Analysis Using Data Mining and Machine Learning Methods. J Med Internet Res 2018 Nov 15;20(11):e12001 [FREE Full text] [doi: 10.2196/12001] [Medline: 30442636]
29. Rahman QA, Janmohamed T, Clarke H, Ritvo P, Heffernan J, Katz J. Interpretability and Class Imbalance in Prediction Models for Pain Volatility in Manage My Pain App Users: Analysis Using Feature Selection and Majority Voting Methods. JMIR Med Inform 2019 Nov 20;7(4):e15601 [FREE Full text] [doi: 10.2196/15601] [Medline: 31746764]
30. Rahman QA, Janmohamed T, Pirbaglou M, Ritvo P, Heffernan JM, Clarke H, et al. Patterns of User Engagement With the Mobile App, Manage My Pain: Results of a Data Mining Investigation. JMIR Mhealth Uhealth 2017 Jul 12;5(7):e96 [FREE Full text] [doi: 10.2196/mhealth.7871] [Medline: 28701291]
31. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. Adm Policy Ment Health 2015 Sep;42(5):533-544 [FREE Full text] [doi: 10.1007/s10488-013-0528-y] [Medline: 24193818]
32. Woo AK. Depression and Anxiety in Pain. Rev Pain 2010 Mar;4(1):8-12 [FREE Full text] [doi: 10.1177/204946371000400103] [Medline: 26527193]
33. Sheng J, Liu S, Wang Y, Cui R, Zhang X. The Link between Depression and Chronic Pain: Neural Mechanisms in the Brain. Neural Plast 2017;2017:9724371 [FREE Full text] [doi: 10.1155/2017/9724371] [Medline: 28706741]

34. Kroenke K, Wu J, Yu Z, Bair MJ, Kean J, Stump T, et al. Patient Health Questionnaire Anxiety and Depression Scale. Psychosom Med 2016;78(6):716-727. [doi: 10.1097/psy.0000000000000322]
35. Kroenke K, Yu Z, Wu J, Kean J, Monahan PO. Operating characteristics of PROMIS four-item depression and anxiety scales in primary care patients with chronic pain. Pain Med 2014 Nov;15(11):1892-1901 [FREE Full text] [doi: 10.1111/pme.12537] [Medline: 25138978]
36. Zhong Q, Gelaye B, Zaslavsky AM, Fann JR, Rondon MB, Sánchez SE, et al. Diagnostic Validity of the Generalized Anxiety Disorder - 7 (GAD-7) among Pregnant Women. PLoS One 2015;10(4):e0125096 [FREE Full text] [doi: 10.1371/journal.pone.0125096] [Medline: 25915929]
37. Seo J, Park S. Validation of the Generalized Anxiety Disorder-7 (GAD-7) and GAD-2 in patients with migraine. J Headache Pain 2015;16:97 [FREE Full text] [doi: 10.1186/s10194-015-0583-8] [Medline: 26596588]
38. Rayner L, Hotopf M, Petkova H, Matcham F, Simpson A, McCracken LM. Depression in patients with chronic pain attending a specialised pain treatment centre: prevalence and impact on health care costs. Pain 2016 Jul;157(7):1472-1479 [FREE Full text] [doi: 10.1097/j.pain.0000000000000542] [Medline: 26963849]
39. Beemster T, van Bennekom C, van Velzen J, Reneman M, Frings-Dresen M. The interpretation of change score of the pain disability index after vocational rehabilitation is baseline dependent. Health Qual Life Outcomes 2018 Sep 14;16(1):182 [FREE Full text] [doi: 10.1186/s12955-018-1000-1] [Medline: 30217206]
40. Quartana PJ, Campbell CM, Edwards RR. Pain catastrophizing: a critical review. Expert Rev Neurother 2009 May;9(5):745-758 [FREE Full text] [doi: 10.1586/ern.09.34] [Medline: 19402782]
41. Sullivan MJL, Thorn B, Rodgers W, Ward LC. Path model of psychological antecedents to pain experience: experimental and clinical findings. Clin J Pain 2004;20(3):164-173. [doi: 10.1097/00002508-200405000-00006] [Medline: 15100592]
42. Perrot S, Lantéri-Minet M. Patients' Global Impression of Change in the management of peripheral neuropathic pain: Clinical relevance and correlations in daily practice. Eur J Pain 2019 Jul;23(6):1117-1128. [doi: 10.1002/ejp.1378] [Medline: 30793414]
43. Kahan M, Mailis-Gagnon A, Tunks E. Canadian guideline for safe and effective use of opioids for chronic non-cancer pain: implications for pain physicians. Pain Res Manag 2011;16(3):157-158 [FREE Full text] [doi: 10.1155/2011/434298] [Medline: 21766064]
44. Fitzmaurice G, Laird N, Ware J. Applied longitudinal Analysis. 2nd ed. Hoboken, NJ: Wiley; 2011.
45. Safi S, Danzer G, Schmailzl KJ. Empirical Research on Acceptance of Digital Technologies in Medicine Among Patients and Healthy Users: Questionnaire Study. JMIR Hum Factors 2019 Nov 29;6(4):e13472 [FREE Full text] [doi: 10.2196/13472] [Medline: 31782741]
46. Greis C, Meier Zürcher C, Djamei V, Moser A, Lautenschlager S, Navarini AA. Unmet digital health service needs in dermatology patients. J Dermatolog Treat 2018 Nov;29(7):643-647. [doi: 10.1080/09546634.2018.1441488] [Medline: 29455570]
47. Slepian PM, Peng M, Janmohamed T, Kotteeswaran Y, Manoo V, Blades AM, et al. Engagement with Manage My Pain mobile health application among patients at the Transitional Pain Service. Digit Health 2020;6:2055207620962297 [FREE Full text] [doi: 10.1177/2055207620962297] [Medline: 33117557]
48. Pratap A, Neto EC, Snyder P, Stepnowsky C, Elhadad N, Grant D, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. NPJ Digit Med 2020;3:21 [FREE Full text] [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
49. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
50. Toussaint A, Hüsing P, Gumz A, Wingenfeld K, Härter M, Schramm E, et al. Sensitivity to change and minimal clinically important difference of the 7-item Generalized Anxiety Disorder Questionnaire (GAD-7). J Affect Disord 2020 Mar 15;265:395-401. [doi: 10.1016/j.jad.2020.01.032] [Medline: 32090765]
51. Sullivan MJL, Bishop SR, Pivik J. The Pain Catastrophizing Scale: Development and validation. Psychol Assess 1995;7(4):524-532. [doi: 10.1037/1040-3590.7.4.524]
52. Lentz TA, Rhon DI, George SZ. Predicting Opioid Use, Increased Health Care Utilization and High Costs for Musculoskeletal Pain: What Factors Mediate Pain Intensity and Disability? J Pain 2020;21(1-2):135-145 [FREE Full text] [doi: 10.1016/j.jpain.2019.06.004] [Medline: 31201989]
53. Martel MO, Wasan AD, Jamison RN, Edwards RR. Catastrophic thinking and increased risk for prescription opioid misuse in patients with chronic pain. Drug Alcohol Depend 2013 Sep 01;132(1-2):335-341 [FREE Full text] [doi: 10.1016/j.drugalcdep.2013.02.034] [Medline: 23618767]
54. Devan H, Hale L, Hempel D, Saipe B, Perry M. What Works and Does Not Work in a Self-Management Intervention for People With Chronic Pain? Qualitative Systematic Review and Meta-Synthesis. Phys Ther 2018 May 01;98(5):381-397. [doi: 10.1093/ptj/pzy029] [Medline: 29669089]
55. Lancaster K, Abuzour A, Khaira M, Mathers A, Chan A, Bui V, et al. The Use and Effects of Electronic Health Tools for Patient Self-Monitoring and Reporting of Outcomes Following Medication Use: Systematic Review. J Med Internet Res 2018 Dec 18;20(12):e294 [FREE Full text] [doi: 10.2196/jmir.9284] [Medline: 30563822]

56. American Psychological Association. Acceptance and Commitment Therapy EST Status for Chronic or Persistent Pain: strong research support. Society of Clinical Psychology (Division 12). 2013. URL: http://www.div12.org/psychological-treatments/disorders/chronic-or-persistent-pain/acceptance-and-commitment-therapy-for-chronic-pain/ [accessed 2015-06-01]
57. Polk KL, Schoendorff B, Webster M, Olaz FO. The Essential Guide to the ACT Matrix: A Step-by-Step Approach to Using the ACT Matrix Model in Clinical Practice. Oakland, CA: New Harbinger Publications; 2016.
58. Andias R, Neto M, Silva AG. The effects of pain neuroscience education and exercise on pain, muscle endurance, catastrophizing and anxiety in adolescents with chronic idiopathic neck pain: a school-based pilot, randomized and controlled study. Physiother Theory Pract 2018 Sep;34(9):682-691. [doi: 10.1080/09593985.2018.1423590] [Medline: 29319386]
59. Oerlemans S, van Cranenburgh O, Herremans P, Spreeuwenberg P, van Dulmen S. Intervening on cognitions and behavior in irritable bowel syndrome: A feasibility trial using PDAs. J Psychosom Res 2011 Mar;70(3):267-277. [doi: 10.1016/j.jpsychores.2010.09.018] [Medline: 21334498]
60. Torbjørnsen A, Ribu L, Rønnevig M, Grøttland A, Helseth S. Users' acceptability of a mobile application for persons with type 2 diabetes: a qualitative study. BMC Health Serv Res 2019 Sep 06;19(1):641 [FREE Full text] [doi: 10.1186/s12913-019-4486-2] [Medline: 31492176]
61. Pilitsis JG, Fahey M, Custozzo A, Chakravarthy K, Capobianco R. Composite Score Is a Better Reflection of Patient Response to Chronic Pain Therapy Compared With Pain Intensity Alone. Neuromodulation 2021 Jan;24(1):68-75. [doi: 10.1111/ner.13212] [Medline: 32592618]
62. Jandoo T. WHO guidance for digital health: What it means for researchers. Digit Health 2020;6:2055207619898984 [FREE Full text] [doi: 10.1177/2055207619898984] [Medline: 31949918]

Abbreviations

GAD-7: Generalized Anxiety Disorder 7-item scale
HCP: health care provider
IFFHT: Iroquois Falls Family Health Team
MMP: Manage My Pain
OME: oral morphine equivalents
PCS: Pain Catastrophizing Scale
PDI: Pain Disability Index
PGIC: Patient Global Impression of Change
PHQ-9: Patient Health Questionnaire 9-item scale
TGH: Toronto General Hospital
TWH: Toronto Western Hospital

Edited by G Eysenbach; submitted 15.12.20; peer-reviewed by Q Rahman, A Silva, R Jamison; comments to author 07.01.21; revised version received 27.01.21; accepted 14.02.21; published 04.03.21.

Please cite as:
Bhatia A, Kara J, Janmohamed T, Prabhu A, Lebovic G, Katz J, Clarke H
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
URL: https://mhealth.jmir.org/2021/3/e26528
doi:10.2196/26528
PMID:33661130

Original Paper

Participants' Engagement and Satisfaction With a Smartphone App Intended to Support Healthy Weight Gain, Diet, and Physical Activity During Pregnancy: Qualitative Study Within the HealthyMoms Trial

Johanna Sandborg[1,2], MSc; Pontus Henriksson[2], PhD; Erica Larsen[2]; Anna-Karin Lindqvist[3], PhD; Stina Rutberg[3], PhD; Emmie Söderström[2], MSc; Ralph Maddison[4], PhD; Marie Löf[1,2,4], PhD

[1]Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden

[2]Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden

[3]Division of Health, Medicine and Rehabilitation, Department of Health, Education and Technology, Luleå University of Technology, Luleå, Sweden

[4]Institute for Physical Activity and Nutrition, School of Exercise and Nutrition Sciences, Deakin University, Burwood, Melbourne, Australia

Corresponding Author:
Johanna Sandborg, MSc
Department of Biosciences and Nutrition
Karolinska Institutet
Group/MLÖ
Huddinge, SE-141 83
Sweden
Phone: 46 733285912
Email: johanna.sandborg@ki.se

Abstract

Background: Excessive gestational weight gain (GWG) is common and associated with negative health outcomes for both mother and child. Mobile health–delivered lifestyle interventions offer the potential to mitigate excessive GWG. The effectiveness of a smartphone app (HealthyMoms) was recently evaluated in a randomized controlled trial. To explore the users' experiences of using the app, a qualitative study within the HealthyMoms trial was performed.

Objective: This qualitative study explored participants' engagement and satisfaction with the 6-month usage of the HealthyMoms app.

Methods: A total of 19 women (mean age: 31.7, SD 4.4 years; mean BMI: 24.6, SD 3.4 kg/m^2; university degree attainment: 13/19, 68%; primiparous: 11/19, 58%) who received the HealthyMoms app in a randomized controlled trial completed semistructured exit interviews. The interviews were audiorecorded and fully transcribed, coded, and analyzed using thematic analysis with an inductive approach.

Results: Thematic analysis revealed a main theme and 2 subthemes. The main theme, "One could suit many: a multifunctional tool to strengthen women's health during pregnancy," and the 2 subthemes, "Factors within and beyond the app influence app engagement" and "Trust, knowledge, and awareness: aspects that can motivate healthy habits," illustrated that a trustworthy and appreciated health and pregnancy app that is easy to use can inspire a healthy lifestyle during pregnancy. The first subtheme discussed how factors within the app (eg, regular updates and feedback) were perceived to motivate both healthy habits and app engagement. Additionally, factors beyond the app were described to both motivate (eg, interest, motivation, and curiosity) and limit (eg, pregnancy-related complications, lack of time) app engagement. The second subtheme reflected important aspects, such as high trustworthiness of the app, increased knowledge, and awareness from using the app, which motivated participants to improve or maintain healthy habits during pregnancy.

Conclusions: The HealthyMoms app was considered a valuable and trustworthy tool to mitigate excessive GWG, with useful features and relevant information to initiate and maintain healthy habits during pregnancy.

Trial Registration: ClinicalTrials.gov NCT03298555; https://clinicaltrials.gov/ct2/show/NCT03298555

International Registered Report Identifier (IRRID): RR2-10.2196/13011

(JMIR Mhealth Uhealth 2021;9(3):e26159) doi:10.2196/26159

KEYWORDS

pregnancy; gestational weight gain; mHealth; telemedicine; digital health; mobile health; eHealth; smartphone intervention; mobile application; smartphone application; engagement; physical activity; exercise; nutrition; diet; qualitative; thematic analysis

Introduction

Excessive Gestational Weight Gain

Excessive gestational weight gain (GWG) is associated with an increased risk of several complications in both mother and child [1,2]. In the short term, these include gestational hypertension, preeclampsia, gestational diabetes mellitus, cesarean delivery, infant macrosomia, preterm birth, and neonatal morbidity and mortality [3,4]. Excessive GWG is also associated with postpartum weight retention and an increased risk of obesity in the offspring, which are indicators of long-term complications [2,4]. Currently, approximately 50% of pregnant women exceed the commonly applied recommendations from the National Academy of Medicine [5] and thereby have an increased risk of different pregnancy-related complications [4,6].

Mobile Health Interventions to Prevent Excessive GWG

While there is strong evidence that a healthy diet, physical activity, or both during pregnancy can reduce the risk of excessive GWG [7,8], conventional in-person programs are time consuming and costly to deliver and have limited reach. Behavior change interventions delivered through mobile phones address some of these issues, as they are easy to access, require fewer resources to deliver once developed, and have greater reach [9-11]. Previous studies have shown positive effects of mobile health (mHealth) programs on weight loss in adults [12], and several pilot studies have shown that mHealth interventions also have the potential to reduce excessive GWG in pregnant women [13-15].

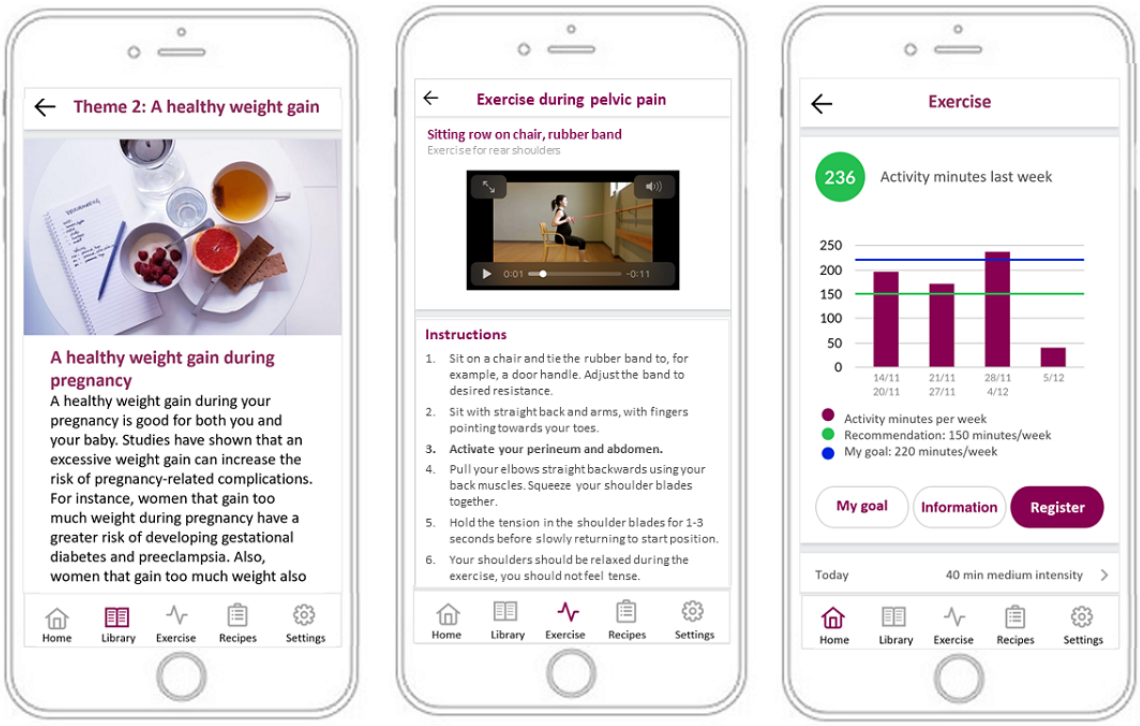
In addition to quantitative evaluations of the effectiveness of mHealth programs on GWG and health behaviors, it is essential to examine user engagement and satisfaction with mHealth programs to facilitate future development, tailoring, and improvements. However, to the best of our knowledge, only 1 study has examined user engagement and satisfaction of an mHealth program for supporting healthy GWG [16]. In brief, in this qualitative pilot study [16], participants (n=13) tested an app (SmartMoms Canada) for 2 to 4 weeks with the aim of evaluating the app prior to testing it in a multicenter study. No previous studies have examined engagement and satisfaction related to long-term usage (ie, 6 months) of an mHealth program to support a healthy weight gain and lifestyle during pregnancy. We have recently developed [17] and evaluated the effectiveness [18] of the smartphone app HealthyMoms, which sought to promote healthy weight gain, diet, and physical activity in pregnancy during 6 months in a randomized controlled trial. This paper reports the findings of a qualitative study that explored the engagement and satisfaction with the HealthyMoms app within the trial.

Methods

Study Design and Recruitment

The HealthyMoms trial (ClinicalTrials.gov NCT03298555) [17] was a 2-arm parallel randomized controlled trial that investigated the effects of a 6-month mHealth intervention (the HealthyMoms app) on GWG, diet, and physical activity. The study was initiated in October 2017, baseline measures (in gestational week 14) were completed in March 2020 (n=305), and follow-up measures (in gestational week 37) were completed in September 2020. After completion of the baseline measurements, participants were randomized (in a 1:1 ratio) to either the control or intervention group. The control group received standard maternity care and the intervention group received the HealthyMoms app (in addition to standard maternity care) for 6 months. The app was developed by researchers with expertise in nutrition, physical activity, pregnancy, and behavior change, and was grounded in social cognitive theory [19] and applied behavior change techniques [20]. HealthyMoms was built around 12 themes and included information and practical tips on diet, physical activity, and weight gain during pregnancy, with a new theme being introduced every other week. The app also included 3 registration features for diet, physical activity, and weight gain. In short, participants were encouraged to register their diet, physical activity, and weight weekly. Registration of diet was based on 5 questions regarding intakes of certain food groups (ie, fruit and vegetables, candy, snacks, sodas, and ice cream and pastries). For physical activity, participants imputed activity minutes as well as intensity of the exercise continuously during the week, and the registrations were summed up for that week. Following registration, participants received feedback based on current national guidelines for diet [21] and physical activity during pregnancy [22]. In the weight registration graph, participants could also view their weight gain compared with their recommended weight gain based on their prepregnancy BMI [5]. Furthermore, the app included an exercise feature (eg, videos and exercise programs), a recipe feature (eg, weekly menus and recipes), and a pregnancy calendar with weekly updates. Figure 1 illustrates 3 of the features (translated screenshots from Swedish to English) in the app. Participants also received push notifications 4 times per week (consisting of tips, encouragement, behavior change strategies, and reminders) and feedback from the registration of diet and physical activity based on national recommendations [21,22]. More detailed description of the rationale, development, and trial design and methods are available in the study protocol [17].

Figure 1. Illustration of 3 app features: information on why a healthy gestational weight gain is important (theme 2), the exercise feature with videos, and the physical activity registration with a goal-setting function and visual feedback in relation to recommendations.



Semistructured Interviews

Women in the intervention group were invited to participate in exit interviews at the follow-up measurement in gestational week 37 (August 2018 to February 2019). The Swedish female researcher (JS) working with the HealthyMoms trial was responsible for recruitment. A total of 20 participants were consecutively asked to participate and all agreed. One participant later declined to participate due to have given birth prior to the scheduled interview. During the final interviews, the research team experienced that no new information was gained and that further interviews were unlikely to provide additional information. We therefore considered that saturation was reached [23] and thus no more women were recruited.

Interviews took place in a separate room at Linköping University, Linköping, Sweden. Participants were interviewed individually and face to face in Swedish by JS. Interviews ranged between 14 and 49 minutes in duration (average time 29 minutes). A semistructured design was used; a set of main questions regarding the app's layout and function and the participants' usage, experiences, and satisfaction using the features in the app were asked, followed by questions tailored to individual responses. The interview guide (Multimedia Appendix 1) was developed by the research team, which has extensive knowledge in pregnancy, nutrition, physical activity, qualitative methodology, and mHealth. Prior to the interview, the participant was given a short introduction to the purpose of the interview and written informed consent was collected. All interviews were audiorecorded and fully transcribed by JS. Audio files were anonymized and kept stored unavailable to unauthorized people. The study was approved by the Regional Ethical Review Board in Linköping, Sweden, on April 24, 2017 (ref No. 2017/112-31), with an amendment on May 4, 2018 (ref No. 2018/262-32). The reporting of this study follows the Consolidated Criteria for Reporting Qualitative Research checklist (Multimedia Appendix 2) [24].

Data Analysis

The semistructured interviews were analyzed using thematic analysis, as described by Braun and Clarke [25]. Data were analyzed using an inductive approach (ie, data-driven) at a semantic level. By using these approaches, there is a lower risk for the analysis to be affected by the preconceptions of the researchers [25]. Interviews were transcribed by JS (nutritionist, PhD student), who had previous experience conducting thematic analysis, and the transcribed texts were then actively read and reread several times by JS and EL (female medical student) to obtain a sense of the overall data. JS and EL separately coded interesting data (ie, text fragments from the transcribed texts) related to the aim of the study (ie, initial codes), and equal attention was given to each data item. The initial codes were then analyzed and sorted into groups to generate themes. Thereafter, any disagreements in the coding or grouping were discussed and resolved before preliminary themes were set. Themes were then reviewed, compared, and contrasted with support from A-KL, SR, ES, and ML, and the final result was derived from thorough discussions between the authors.

Results

Participants

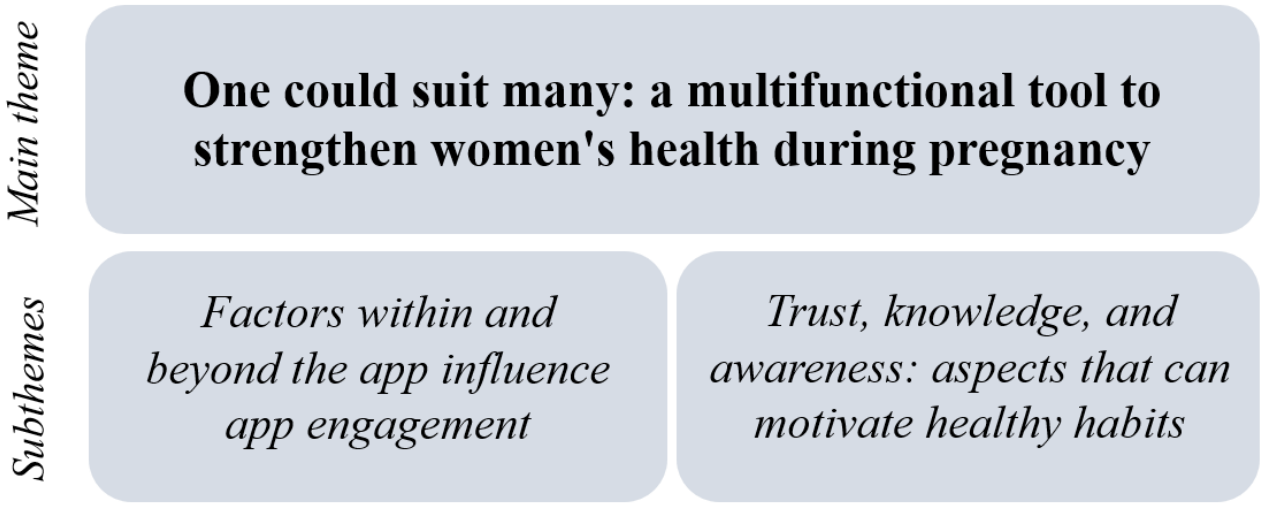
The 19 interviewed women were representative of the whole intervention group (n=152) in terms of baseline characteristics such as average age (31.7, SD 4.4 years vs 31.4, SD 4.3 years), BMI (24.6, SD 3.4 kg/m^2 vs 24.7, SD 4.3 kg/m^2), educational attainment (13/19, 68% vs 115/152, 76% had a university degree), and parity (11/19, 58% vs 86/152, 57% were primiparous). Also, the 19 women participating in these

interviews reported similar usage and satisfaction by means of a questionnaire after finalizing the intervention compared with women in the intervention group that completed the follow-up (n=134) [18]. For instance, they reported app usage comparable to the whole intervention group (15/19, 79% vs 111/134, 83% reported having used the app ≥1 time per week), and the majority reported that they agreed with the statement that they were satisfied with the app (18/19, 95% vs 122/134, 91%). Further, participants in the interview group and the whole intervention group agreed to a comparable extent with the statement that the app was a good support for healthy GWG (11/19, 58% vs 91/134, 68%, respectively).

Themes From the Thematic Analysis

One major theme and 2 subthemes were derived from the thematic analysis and are presented in Figure 2.

Figure 2. The themes identified in the thematic analysis.



One Could Suit Many: A Multifunctional Tool to Strengthen Women's Health During Pregnancy

The themes illustrated that a trustworthy and appreciated health and pregnancy app that was easy to use could inspire healthy habits during pregnancy. Overall, the app was appreciated but used in different ways. The first subtheme discussed factors within the app (eg, regular updates, feedback) and beyond the app (eg, motivation, interest, and the pregnancy itself) that influenced app engagement. The second subtheme reflected that trust, increased awareness, and knowledge were important and appreciated aspects in the app, as these motivated participants to improve or maintain healthy habits.

Factors Within and Beyond the App Influence App Engagement

Engagement with the app varied among participants, and several factors were described to influence app engagement. For instance, regular updates (with new app content and information) and push notifications were described as positive, as they sparked interest and increased usage. However, for some women, positive feedback on diet registrations actually decreased motivation if goals were repeatedly reached, which indicated no room for further improvement. Furthermore, when goals were not met, some women reported that feedback suggesting the need for improvement led to feelings of guilt and negative emotions, resulting in discontinued registration.

> *Too much time does not pass before you enter the app again, if it had I do not think I would have registered my own data as regularly, because otherwise I think you kind of would have forgotten that. So, I felt that I used the app more because of the push notifications.* [Participant 5]

Furthermore, the risk of becoming too fixated with weight and diet was mentioned as a reason for not engaging in these registration features. Compared with the other registration features, participants described being less motivated to use the diet registration because of difficulties remembering weekly food intakes and perceiving feedback as ambiguous. A more detailed diet registration was suggested as a potential improvement; however, it was also described that such registration could be too time-consuming and potentially put an unhealthy amount of focus on diet. Registration of physical activity was also described as confusing and choosing the appropriate intensity level was difficult. Thus, a suggested improvement was the option to choose the activity (eg, running, swimming, walking) instead of intensity level. Other suggested improvements were an inbuilt pedometer or the possibility of transferring data (manually or automatically) from other apps.

> *I did not quite understand the diet registration graph and because of that I could not tell if my diet improved or got worse. If it would have been easier to understand I think I probably would have been more motivated to continue to register my diet.* [Participant 2]

Several other factors beyond the app impacted usage. First, app engagement was described to be influenced by both lifestyle and prior knowledge as well as the experienced need of the features in the app (eg, need for motivation to exercise, to track GWG, or tips and information on how to eat healthy). For example, low or nonexistent usage of the diet registration could be explained by the lack of a need or motivation to make dietary

changes and thus register diet. Furthermore, personal interests influenced app usage, especially the use of the recipe and exercise-related features.

> *The exercise functions were the best parts of the app, but that is also because I like to move and to be active, so that is probably the part that I have used the most.* [Participant 15]

Moreover, it was described that the usefulness of the app was greater when participants were motivated and experienced the need to change behavior. For instance, having gained too much weight during a previous pregnancy was described to motivate app usage. On the same note, parity was described to impact engagement and usefulness of the app, as women expecting their second child experienced less need of the pregnancy-related features compared with first-time mothers. Additionally, life situation (eg, number of children, workload) was mentioned to potentially impact app engagement, as it determined the available time to spend on an app. Nevertheless, both primiparous and multiparous women were satisfied with the app and found it to be useful.

> *I have not had the time to be active in any type of app or anything like that so I think that it is more those things [work, other children] that have made me feel that I have not had the time as I did during my first pregnancy when I literally read everything, so it has been a bit different this time.* [Participant 14]

Participants also reported that their engagement with the app changed during the course of their pregnancy. Curiosity about the app and the need of pregnancy-related information were higher in early pregnancy, but as women established new, healthier habits and perceived an increased sense of security as pregnancy progressed, app usage declined in late pregnancy. App engagement was sometimes negatively influenced by pregnancy complications (eg, nausea or pelvic pain), which could inhibit the ability and motivation to maintain a healthy lifestyle and consequently use the self-monitoring features in the app. Finally, initial higher app usage was explained mainly by participants having more energy and motivation in the beginning and middle of the pregnancy. In contrast, one participant described that the pregnancy initially felt more surreal, and as pregnancy progressed, engaging in healthy habits felt more important, and consequently app usage increased.

> *I was more curious of the app in the beginning, more curious of all the content, and then as time passed by I sort of only used the functions that I liked the most.* [Participant 6]

> *I was more committed [to the app] towards the end because then I kind of felt that it mattered more since I started to feel more pregnant.* [Participant 5]

Trust, Knowledge, and Awareness: Aspects That Can Motivate Healthy Habits

Participants stated that they appreciated the app because it had an appealing layout, was easy to use, had no technical shortcomings, and was perceived as a trustworthy source of information, as it was developed by experts. Furthermore, one participant described that the credibility of the app was strengthened by the fact that it was noncommercial. Another participant appreciated that the app content was in line with Swedish maternity care, which made the app more relevant and useful compared with other foreign pregnancy apps.

> *I've trusted this app considering the people behind it. If I am to be source-critical, I've found it very nice to have got it [my preconceptions] confirmed and I feel that I can trust that what I'm reading in the app is true.* [Participant 9]

Moreover, the app provided necessary information (eg, in the pregnancy calendar and themes). Thus, participants did not have to look for information elsewhere, which was considered valuable. The multifunctionality of the app, inclusion of registration features, and focus on both general and pregnancy-related health were appreciated and described as rare in other pregnancy apps. The wide range of features were also described to increase its usefulness, and thus the app could benefit many different women. Participants stated that HealthyMoms confirmed and increased their knowledge about healthy weight gain, diet, and physical activity during pregnancy, and increased awareness of how lifestyle could impact both maternal and fetal health was described as motivating. The registration features enabled self-evaluation and were considered a good support for changing or maintaining habits as well as supporting a healthy GWG. Participants reported that app usage had a positive impact on their diet and physical activity regardless of their habits prior to receiving the app.

> *I've become more motivated to eat well and exercise for my own health and because it [a healthy lifestyle] facilitates my pregnancy and also that it is good for the baby.* [Participants 19]

Additionally, information on the importance of healthy GWG was appreciated, and one participant described that it is not usually discussed at midwife appointments. However, the recommended GWG (based on the self-monitoring of weight in the app) was described as stricter than information received from the midwife, which brought up the importance of consistent information. Additionally, exceeding the GWG recommendations could cause feelings of anxiety, discouragement, and frustration.

> *I thought that [the recommended weight gain in the graph] was good because I do not feel that you get that information elsewhere, on how much weight you should gain, or that is something that is not spoken of at the midwife visits, well you talk about weight but not so much from that perspective.* [Participant 12]

Overall, self-monitoring was seen as motivating, as it increased awareness of GWG and dietary and exercise habits. On the same note, push notifications reminded participants of the importance of a healthy lifestyle, which was motivating; however, these could also be perceived as annoying when received at an inconvenient time (eg, receiving a message on the importance of a healthy diet right after having eaten something unhealthy). Moreover, self-monitoring and goal setting made the app feel personal, which was considered important. Participants

described, however, that the app could be improved by adding additional personalized features, such as a calendar (to schedule, for example, exercise or midwife visits), tailored push notifications (eg, feedback related to registration), and challenges (eg, minor assignments such as going for a walk or having a piece of fruit).

Moreover, additional functions and information that were described as desirable included a larger focus on mental health, a sharing and network function, and more inclusion of the partner. Furthermore, to prevent the loss of healthy habits in early pregnancy, participants wished for earlier access to the app. A prolonged version covering the postpartum period to support a healthy lifestyle and the inclusion of baby-related advice were described as additional improvements.

> *I think it is important that it [the app] continues to be personal, that you feel that it is adapted to my premises and my starting weight, my BMI, and that you get to self-register as an individual and get data from that. I think that is the most important feature and what also perhaps separates this app from all other apps.* [Participant 17]

Discussion

Principal Results

Overall, this study demonstrated various levels of engagement and overall high satisfaction with the HealthyMoms app, and women perceived that it inspired healthy habits during pregnancy. Engagement with HealthyMoms was influenced by the app itself (eg, regular updates, feedback, and push notifications) and factors beyond the app (eg, interest, need, the pregnancy, and time constraints). Important and appreciated aspects of the app were trustworthiness, information, and features that increased awareness, knowledge, and motivation.

Comparison With Prior Work

To the best of our knowledge, this is the first study to explore participants' engagement and satisfaction related to long-term usage (ie, 6 months) of an mHealth app intended to support healthy weight gain and lifestyle during pregnancy. This study identified several factors both within and beyond the app that influenced engagement with the HealthyMoms app. Regular updates, push notifications, and feedback from the self-monitoring features sparked interest and positively influenced app engagement. Similarly, a study in a nonpregnant population found reminders to be positive; however, push notifications need to be carefully constructed in terms of timing and frequency, as they otherwise might be ignored [26]. This is similar to our study, where some participants described push notifications to be annoying when received at a bad time.

Although factors within the app influenced usage, factors beyond the app seemed to influence usage to a greater extent. Similar to a previous study exploring the acceptability of a multicomponent intervention in 9 Australian pregnant women [27], we also found that life situation and available time limited app engagement. Pregnancy itself was also described to guide app engagement; usage of the HealthyMoms app tended to peak in early pregnancy and then declined, which was explained by higher curiosity in the beginning. The same usage pattern has been seen in a previous study in Chinese pregnant women investigating the use of smartphone apps in general [28]. However, we also found that increased usage could occur during the final period due to feeling more pregnant and thus being more motivated to engage in healthy habits. Similar to previous studies investigating app usage in nonpregnant populations [26,29], app engagement was further described to be influenced by motivation, and usage could decrease with time due to the establishment of new habits. The usage of the HealthyMoms app also depended on pregnancy-related complications. These findings are similar to previous results by Willcox et al [27], in which participants described pregnancy-related physical ailments, such as back pain and morning sickness, as barriers to maintaining a healthy lifestyle. Consideration of these factors could be important for the development of future health and pregnancy apps.

Our findings show that the HealthyMoms app was perceived as trustworthy and a reliable source of information, as it contained evidence-based information and was developed by experts. This finding is supported by previous data on pregnant women's app usage in general that have shown that unreliable or uncertain information can cause feelings of anxiety [28,30]. Furthermore, the importance of engagement from health care professionals and institutions in the development of this type of app have previously been emphasized [30,31]. Both users and health care professionals have been shown to prefer pregnancy apps that contain information related to their local health care context [30]. Likewise, participants in our study appreciated that the app was context specific and relevant to Swedish maternity care.

In line with previous studies [28,30-32], women in this study valued the HealthyMoms app for the wide range of features and its focus on both pregnancy and health. The opportunity to self-monitor was appreciated, and it increased awareness of dietary and exercise habits as well as GWG, which motivated a higher usage of the app and aided in establishing healthier habits. This is also in line with previous findings, which have reported that self-monitoring of weight during pregnancy could be perceived as helpful and motivating to stay within the recommendations for GWG [16]. However, the feature for weight registration could also be perceived as stressful and cause anxiety when exceeding the recommendations. In this study, participants also expressed that GWG in relation to the recommended weight gain is not usually spoken of in maternity care. Indeed, both midwives and pregnant women themselves have expressed challenges and stigma related to discussing GWG in maternity care [33-35]. Since GWG is important to address, an app with a carefully designed weight registration can fill an important gap.

Overall, participants were satisfied with the HealthyMoms app. However, improvements for future versions were identified. One function described as desirable was a sharing and network feature, which previously has been reported as an appreciated function by pregnant women [16,30,31]. Although the personalized aspects of the app (ie, the self-monitoring features) were highly appreciated, making goals and feedback more individualized was a suggested improvement. Likewise,

individualization has been found to be important in previous studies in both pregnant [16] and nonpregnant populations [26]. Similar to a pilot study in Canadian women that evaluated an app (SmartMoms Canada) after a test period of 2 to 4 weeks [16], the participants in this study wished for a larger focus on mental health. This could be important to consider in future versions of the HealthyMoms app or other apps targeting pregnant women. Participants in our study wished for earlier access to the app in order to prevent the loss of healthy habits. They also expressed a desire for a continued version of the app after the baby's arrival in order to support a healthy lifestyle after birth and for the app to include baby-related advice, which has been indicated as a requested feature in other populations of pregnant women [30,33]. Potentially, future app versions should be more comprehensive and span from early pregnancy into infancy.

In this study, the diet registration was perceived as difficult to use and should be improved in the future. A more detailed registration was suggested, but that could potentially lead to an unhealthy preoccupation with food. Moreover, our results suggest that self-monitoring of physical activity should include more detailed registrations, such as types of activity. Furthermore, an inbuilt pedometer or the opportunity to transfer data from other apps was expressed as a desirable addition. In comparison, Halili et al [16] combined an app (SmartMoms Canada) with a fitness tracker, which was appreciated by the pregnant women after a short test period. Ease of use and simplicity have also been found to be important in overcoming barriers (eg, lack of time) to app usage in nonpregnant populations [26]. Taken together, these suggested improvements could be important to consider for future version of the HealthyMoms app as well as future pregnancy and health apps.

Strengths and Limitations

In this study, we have presented a qualitative evaluation of the engagement and satisfaction with HealthyMoms. This type of evaluation is important to guide future interventions, the development of other pregnancy and health apps, and future versions of HealthyMoms. Qualitative data also provide a more detailed evaluation of how women engaged with the program compared with quantitative measures only.

A strength of this study is the inclusion of 2 people (JS and EL) to undertake the thematic analysis, which allowed for constant validation during the process. The analysis was further strengthened by having 2 perspectives on the data, one from the person who conducted the interviews (JS) and one from someone who did not meet the participating women (EL); thus, comprehension of the data was discussed and validated with a low risk of biased interpretations.

This study was limited by the fact that only 19 out of the 134 participants in the intervention group participated in the interviews. However, importantly, the sample size was sufficient to obtain a broad and rich variety of experiences, the interviews demonstrated various engagement with the app as well as varying attitudes, and enough data were collected to achieve perceived saturation [23]. Additionally, the participants were representative of the intervention group. The participants had a higher education level compared with the general population, but this was similar to the education level in the entire group.

Implications and Clinical Relevance

The HealthyMoms app was well received by participants and the multifunctionality of the app, including the self-monitoring features, were appreciated. Our results also provide valuable information that can be used to further iterate and improve the HealthyMoms app as well as guide the development of future pregnancy apps intended for long-term usage. Such minor improvements include more personalized features, a larger focus on mental health, and an inbuilt pedometer. Given that midwives may have little time to fully address and support healthy GWG in maternity care and due to the sensitive nature of weight gain, this app has the potential to fill an important gap. Considering that reducing the burden of excessive GWG has significant and important health implications for both mother and child [1,4], access to low-cost and scalable tools to promote healthy GWG is essential. The usage of pregnancy apps among women in developed countries is extensive [28,31,36], and a health and pregnancy app could therefore potentially reach many women.

Conclusion

This study contributes novel information about pregnant women's experiences using a health and pregnancy app intended to support a healthy GWG throughout pregnancy. The results showed that the HealthyMoms app was appreciated and is likely to benefit pregnant women. Several factors with both positive (eg, push notifications, regular updates, interest) and negative (eg, time constraints, pregnancy-related complications) impact on app engagement were observed. Trust, knowledge, and awareness were described as important aspects to motivate healthy habits. Taken together, our results showed that the app was considered a valuable support and could aid in establishing and maintaining healthy habits during pregnancy, which is important to prevent excessive GWG. Finally, our findings provide valuable information for the development of future pregnancy apps intended to support a healthy lifestyle.

Acknowledgments

The authors would like to thank the participating women. The following funders provided support for this study: the Swedish Research Council (2016-01147 [ML]); the Swedish Research Council for Health, Working Life and Welfare (Forte, 2017-00088 [PH]; 2018-01410 [ML]); the Strategic Research Area Health Care Science, Karolinska Institutet and Umeå University (PH); the Swedish Society of Medicine (PH); Karolinska Institutet (PH); Lions Forskningsfond (PH); and ALF Grants, Region Östergötland (ML). The funders had no role in the design, conduct, data management, or interpretation of the analysis or in the decision to submit the manuscript for publication.

Authors' Contributions

ML designed the study together with JS, RM, and PH. JS recruited the participants, performed all interviews, and transcribed the interviews. JS and EL conducted the thematic analysis with support from A-KL, SR, ES, and ML. JS wrote the manuscript, which was critically reviewed by PH, EL, A-KL, SR, ES, RM, and ML. All authors approved of the final draft of the submitted manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The interview guide.
[DOCX File , 29 KB - mhealth_v9i3e26159_app1.docx]

Multimedia Appendix 2
The Consolidated Criteria for Reporting Qualitative Research checklist.
[PDF File (Adobe PDF File), 481 KB - mhealth_v9i3e26159_app2.pdf]

References

1. Santos S, Voerman E, Amiano P, Barros H, Beilin LJ, Bergström A, et al. Impact of maternal body mass index and gestational weight gain on pregnancy complications: an individual participant data meta-analysis of European, North American and Australian cohorts. BJOG 2019 Jul;126(8):984-995 [FREE Full text] [doi: 10.1111/1471-0528.15661] [Medline: 30786138]
2. LifeCycle Project-Maternal Obesity and Childhood Outcomes Study Group, Voerman E, Santos S, Inskip H, Amiano P, Barros H, et al. Association of Gestational Weight Gain With Adverse Maternal and Infant Outcomes. JAMA 2019 May 07;321(17):1702-1715 [FREE Full text] [doi: 10.1001/jama.2019.3820] [Medline: 31063572]
3. Simko M, Totka A, Vondrova D, Samohyl M, Jurkovicova J, Trnka M, et al. Maternal Body Mass Index and Gestational Weight Gain and Their Association with Pregnancy Complications and Perinatal Conditions. Int J Environ Res Public Health 2019 May 17;16(10):1751 [FREE Full text] [doi: 10.3390/ijerph16101751] [Medline: 31108864]
4. Goldstein RF, Abell SK, Ranasinha S, Misso M, Boyle JA, Black MH, et al. Association of Gestational Weight Gain With Maternal and Infant Outcomes: A Systematic Review and Meta-analysis. JAMA 2017 Jun 06;317(21):2207-2225 [FREE Full text] [doi: 10.1001/jama.2017.3635] [Medline: 28586887]
5. Rasmussen K, Yaktine A. Weight Gain During Pregnancy: Reexamining the Guidelines. Washington, DC: The National Academies Press (US); 2009.
6. Henriksson P, Sandborg J, Blomberg M, Nowicka P, Petersson K, Bendtsen M, et al. Body mass index and gestational weight gain in migrant women by birth regions compared with Swedish-born women: A registry linkage study of 0.5 million pregnancies. PLoS One 2020;15(10):e0241319 [FREE Full text] [doi: 10.1371/journal.pone.0241319] [Medline: 33119672]
7. Muktabhant B, Lawrie TA, Lumbiganon P, Laopaiboon M. Diet or exercise, or both, for preventing excessive weight gain in pregnancy. Cochrane Database Syst Rev 2015 Jun 15;6:CD007145. [doi: 10.1002/14651858.CD007145.pub3] [Medline: 26068707]
8. Watson ED, Macaulay S, Lamont K, Gradidge PJL, Pretorius S, Crowther NJ, et al. The effect of lifestyle interventions on maternal body composition during pregnancy in developing countries: a systematic review. Cardiovasc J Afr 2017;28(6):397-403 [FREE Full text] [doi: 10.5830/CVJA-2017-003] [Medline: 28906530]
9. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
10. Stephenson A, McDonough SM, Murphy MH, Nugent CD, Mair JL. Using computer, mobile and wearable technology enhanced interventions to reduce sedentary behaviour: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2017 Aug 11;14(1):105 [FREE Full text] [doi: 10.1186/s12966-017-0561-4] [Medline: 28800736]
11. McKay FH, Wright A, Shill J, Stephens H, Uccellini M. Using Health and Well-Being Apps for Behavior Change: A Systematic Search and Rating of Apps. JMIR Mhealth Uhealth 2019 Jul 04;7(7):e11926 [FREE Full text] [doi: 10.2196/11926] [Medline: 31274112]
12. 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]
13. Redman LM, Gilmore LA, Breaux J, Thomas DM, Elkind-Hirsch K, Stewart T, et al. Effectiveness of SmartMoms, a Novel eHealth Intervention for Management of Gestational Weight Gain: Randomized Controlled Pilot Trial. JMIR Mhealth Uhealth 2017 Sep 13;5(9):e133 [FREE Full text] [doi: 10.2196/mhealth.8228] [Medline: 28903892]
14. Willcox JC, Wilkinson SA, Lappas M, Ball K, Crawford D, McCarthy EA, et al. A mobile health intervention promoting healthy gestational weight gain for women entering pregnancy at a high body mass index: the txt4two pilot randomised controlled trial. BJOG 2017 Oct;124(11):1718-1728. [doi: 10.1111/1471-0528.14552] [Medline: 28220604]

15. Overdijkink SB, Velu AV, Rosman AN, van Beukering MD, Kok M, Steegers-Theunissen RP. The Usability and Effectiveness of Mobile Health Technology-Based Lifestyle and Medical Intervention Apps Supporting Health Care During Pregnancy: Systematic Review. JMIR Mhealth Uhealth 2018 Apr 24;6(4):e109 [FREE Full text] [doi: 10.2196/mhealth.8834] [Medline: 29691216]
16. Halili L, Liu R, Hutchinson KA, Semeniuk K, Redman LM, Adamo KB. Development and pilot evaluation of a pregnancy-specific mobile health tool: a qualitative investigation of SmartMoms Canada. BMC Med Inform Decis Mak 2018 Nov 12;18(1):95 [FREE Full text] [doi: 10.1186/s12911-018-0705-8] [Medline: 30419896]
17. Henriksson P, Sandborg J, Blomberg M, Alexandrou C, Maddison R, Silfvernagel K, et al. A Smartphone App to Promote Healthy Weight Gain, Diet, and Physical Activity During Pregnancy (HealthyMoms): Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2019 Mar 01;8(3):e13011 [FREE Full text] [doi: 10.2196/13011] [Medline: 30821695]
18. Sandborg J, Söderström E, Henriksson P, Bendtsen M, Henström M, Leppänen MH, et al. Effectiveness of a Smartphone App to Promote Healthy Weight Gain, Diet, and Physical Activity During Pregnancy (HealthyMoms): Randomized Controlled Trial [forthcoming]. JMIR mHealth and uHealth 2021. [doi: 10.2196/26091]
19. Bandura A. Human agency in social cognitive theory. Am Psychol 1989 Sep;44(9):1175-1184. [doi: 10.1037/0003-066x.44.9.1175] [Medline: 2782727]
20. 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]
21. Nordic Council of Ministers. Nordic Nutrition Recommendations 2012. 2014. URL: https://www.livsmedelsverket.se/globalassets/publikationsdatabas/andra-sprak/nordic-nutrition-recommendations-2012.pdf?AspxAutoDetectCookieSupport=1 [accessed 2020-10-27]
22. Josefsson A, Haakstad LA, Bö K. Rekommendationer om fysisk aktivitet vid graviditet. FYSS-kapitel Fysisk Aktivitet Vid Graviditet. 2016. URL: http://www.fyss.se/wp-content/uploads/2017/09/FYSS-kapitel_FA_Graviditet_FINAL_2016-12.pdf [accessed 2020-10-28]
23. Faulkner SL, Trotter SP. Data Saturation. In: The International Encyclopedia of Communication Research Methods. Hoboken, NJ: Wiley; Nov 07, 2017:1-2.
24. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
25. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
26. 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]
27. Willcox JC, Chai D, Beilin LJ, Prescott SL, Silva D, Neppe C, et al. Evaluating Engagement in a Digital and Dietetic Intervention Promoting Healthy Weight Gain in Pregnancy: Mixed Methods Study. J Med Internet Res 2020 Jun 26;22(6):e17845 [FREE Full text] [doi: 10.2196/17845] [Medline: 32442153]
28. Wang N, Deng Z, Wen LM, Ding Y, He G. Understanding the Use of Smartphone Apps for Health Information Among Pregnant Chinese Women: Mixed Methods Study. JMIR Mhealth Uhealth 2019 Jun 18;7(6):e12631 [FREE Full text] [doi: 10.2196/12631] [Medline: 31215516]
29. 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]
30. Hughson JP, Daly JO, Woodward-Kron R, Hajek J, Story D. The Rise of Pregnancy Apps and the Implications for Culturally and Linguistically Diverse Women: Narrative Review. JMIR Mhealth Uhealth 2018 Nov 16;6(11):e189 [FREE Full text] [doi: 10.2196/mhealth.9119] [Medline: 30446483]
31. Lupton D. The use and value of digital media for information about pregnancy and early motherhood: a focus group study. BMC Pregnancy Childbirth 2016 Jul 19;16(1):171 [FREE Full text] [doi: 10.1186/s12884-016-0971-3] [Medline: 27435182]
32. Soltani H, Duxbury AMS, Arden MA, Dearden A, Furness PJ, Garland C. Maternal Obesity Management Using Mobile Technology: A Feasibility Study to Evaluate a Text Messaging Based Complex Intervention during Pregnancy. J Obes 2015;2015:814830 [FREE Full text] [doi: 10.1155/2015/814830] [Medline: 25960889]
33. Furness PJ, McSeveny K, Arden MA, Garland C, Dearden AM, Soltani H. Maternal obesity support services: a qualitative study of the perspectives of women and midwives. BMC Pregnancy Childbirth 2011 Oct 08;11:69 [FREE Full text] [doi: 10.1186/1471-2393-11-69] [Medline: 21982306]
34. Flannery C, McHugh S, Kenny LC, O'Riordan MN, McAuliffe FM, Bradley C, et al. Exploring obstetricians', midwives' and general practitioners' approach to weight management in pregnant women with a BMI ≥25 kg/m: a qualitative study. BMJ Open 2019 Jan 28;9(1):e024808 [FREE Full text] [doi: 10.1136/bmjopen-2018-024808] [Medline: 30696684]
35. Willcox JC, Campbell KJ, van der Pligt P, Hoban E, Pidd D, Wilkinson S. Excess gestational weight gain: an exploration of midwives' views and practice. BMC Pregnancy Childbirth 2012 Sep 27;12:102 [FREE Full text] [doi: 10.1186/1471-2393-12-102] [Medline: 23013446]

36. O'Higgins A, Murphy OC, Egan A, Mullaney L, Sheehan S, Turner MJ. The use of digital media by women using the maternity services in a developed country. Ir Med J 2014;107(10):313-315. [Medline: 25556255]

Abbreviations

GWG: gestational weight gain
mHealth: mobile health

Edited by G Eysenbach; submitted 30.11.20; peer-reviewed by A Beleigoli, R Lim, N Broskey; comments to author 21.12.20; revised version received 19.01.21; accepted 30.01.21; published 05.03.21.

Please cite as:
Sandborg J, Henriksson P, Larsen E, Lindqvist AK, Rutberg S, Söderström E, Maddison R, Löf M
Participants' Engagement and Satisfaction With a Smartphone App Intended to Support Healthy Weight Gain, Diet, and Physical Activity During Pregnancy: Qualitative Study Within the HealthyMoms Trial
JMIR Mhealth Uhealth 2021;9(3):e26159
URL: https://mhealth.jmir.org/2021/3/e26159
doi:10.2196/26159
PMID:33666554

Original Paper

Effectiveness of a Smartphone App to Promote Healthy Weight Gain, Diet, and Physical Activity During Pregnancy (HealthyMoms): Randomized Controlled Trial

Johanna Sandborg[1,2], MSc; Emmie Söderström[2], MSc; Pontus Henriksson[2], PhD; Marcus Bendtsen[2], PhD; Maria Henström[1], PhD; Marja H Leppänen[1,3,4], PhD; Ralph Maddison[5], PhD; Jairo H Migueles[2,6], PhD; Marie Blomberg[7,8], PhD; Marie Löf[1,2,5], PhD

[1]Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden

[2]Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden

[3]Folkhälsan Research Center, Helsinki, Finland

[4]Faculty of Medicine, University of Helsinki, Helsinki, Finland

[5]Institute for Physical Activity and Nutrition, Deakin University, Burwood, Australia

[6]PROFITH "PROmoting FITness and Health through physical activity" research group, Department of Physical Education and Sports, Faculty of Sport Sciences, Research Institute of Sport and Health, University of Granada, Granada, Spain

[7]Department of Obstetrics and Gynecology, Linköping University, Linköping, Sweden

[8]Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden

Corresponding Author:
Johanna Sandborg, MSc
Department of Biosciences and Nutrition
Karolinska Institutet
Group/MLÖ
Huddinge, SE-141 83
Sweden
Phone: 46 733285912
Email: johanna.sandborg@ki.se

Abstract

Background: Excessive gestational weight gain (GWG) during pregnancy is a major public health concern associated with negative health outcomes for both mother and child. Scalable interventions are needed, and digital interventions have the potential to reach many women and promote healthy GWG. Most previous studies of digital interventions have been small pilot studies or have not included women from all BMI categories. We therefore examined the effectiveness of a smartphone app in a large sample (n=305) covering all BMI categories.

Objective: To investigate the effectiveness of a 6-month intervention (the HealthyMoms app) on GWG, body fatness, dietary habits, moderate-to-vigorous physical activity (MVPA), glycemia, and insulin resistance in comparison to standard maternity care.

Methods: A 2-arm parallel randomized controlled trial was conducted. Women in early pregnancy at maternity clinics in Östergötland, Sweden, were recruited. Eligible women who provided written informed consent completed baseline measures, before being randomized in a 1:1 ratio to either an intervention (n=152) or control group (n=153). The control group received standard maternity care while the intervention group received the HealthyMoms smartphone app for 6 months (which includes multiple features, eg, information; push notifications; self-monitoring; and feedback features for GWG, diet, and physical activity) in addition to standard care. Outcome measures were assessed at Linköping University Hospital at baseline (mean 13.9 [SD 0.7] gestational weeks) and follow-up (mean 36.4 [SD 0.4] gestational weeks). The primary outcome was GWG and secondary outcomes were body fatness (Bod Pod), dietary habits (Swedish Healthy Eating Index) using the web-based 3-day dietary record Riksmaten FLEX, MVPA using the ActiGraph wGT3x-BT accelerometer, glycemia, and insulin resistance.

Results: Overall, we found no statistically significant effect on GWG (*P*=.62); however, the data indicate that the effect of the intervention differed by pre-pregnancy BMI, as women with overweight and obesity before pregnancy gained less weight in the intervention group as compared with the control group in the imputed analyses (–1.33 kg; 95% CI –2.92 to 0.26; *P*=.10) and

completers-only analyses (–1.67 kg; 95% CI –3.26 to –0.09; *P*=.031]). Bayesian analyses showed that there was a 99% probability of any intervention effect on GWG among women with overweight and obesity, and an 81% probability that this effect was over 1 kg. The intervention group had higher scores for the Swedish Healthy Eating Index at follow-up than the control group (0.27; 95% CI 0.05-0.50; *P*=.017). We observed no statistically significant differences in body fatness, MVPA, glycemia, and insulin resistance between the intervention and control group at follow up (*P*≥.21).

Conclusions: Although we found no overall effect on GWG, our results demonstrate the potential of a smartphone app (HealthyMoms) to promote healthy dietary behaviors as well as to decrease weight gain during pregnancy in women with overweight and obesity.

Trial Registration: ClinicalTrials.gov NCT03298555; https://clinicaltrials.gov/ct2/show/NCT03298555

International Registered Report Identifier (IRRID): RR2-10.2196/13011

(JMIR Mhealth Uhealth 2021;9(3):e26091) doi:10.2196/26091

KEYWORDS

gestational weight gain; physical activity; diet; pregnancy; mHealth; smartphone app; mobile phone app; telemedicine; randomized controlled trial

Introduction

Background

Excessive gestational weight gain (GWG) is a major public health problem [1,2]. In the United States and Europe, around 50% of pregnant women exceed the recommended GWG provided by the National Academy of Medicine [1] with similar data from Sweden [3,4]. Excessive GWG is associated with increased risk of cesarean delivery, gestational diabetes, pre-eclampsia, and obesity in both mother and child [2,5]. Furthermore, previous studies have shown that approximately 20% of American and European women gain less weight than recommended during pregnancy [1], which is associated with complications such as an increased risk of low birth weight and preterm birth [6]. Thus, effective and evidence-based strategies to promote a healthy GWG are of great importance.

Traditional interventions (eg, face-to-face counseling and supervised exercise sessions) to reduce the risk of excessive GWG have been reported to be successful [7-9]. For example, a Cochrane review [9] found that traditional interventions focusing on diet, exercise, or both have been found to reduce the risk of excessive GWG by 20%. Correspondingly, a recent systematic review and meta-analysis [10] found that interventions aiming to improve diet and physical activity behaviors have shown to reduce GWG by 1.81 kg (95% CI –3.47 to –0.16; 21 studies, n=6920) in pregnant women with overweight and obesity [10]. However, traditional interventions are generally resource heavy and rely considerably on health care staff. Furthermore, they are costly and lack the ability to reach large numbers of women.

In the last decade, the use of digital technologies (eg, mobile Health [mHealth]) to deliver lifestyle interventions has increased. In comparison to traditional interventions, mHealth interventions have the advantages of being more cost-effective and accessible [11] and may reduce burden on health care staff. Accumulating evidence indicates that interventions delivered using this technology may promote weight loss [12] and increase physical activity in adults [13]. Furthermore, a recent mHealth pilot study (n=54) in pregnant women with overweight and obesity found that a lower proportion in the intervention group exceeded the recommended GWG compared with the control group who received usual care (58% vs 85%) [14]. However, as highlighted in a recent review [15], mHealth interventions focusing on healthy GWG have generally been pilot studies with small sample sizes (eg, [14,16-18]), or have not included all BMI categories (eg, [19-22]). Moreover, to the best of our knowledge, no previous study has investigated the effectiveness of a behavior change program delivered solely through a smartphone app on GWG in women covering all BMI categories.

Aim

The aim of this randomized controlled trial was to investigate the effectiveness of the 6-month intervention (the HealthyMoms app) on GWG (primary outcome), body fatness, dietary habits (Swedish Healthy Eating Index), moderate-to-vigorous physical activity (MVPA), glycemia, and insulin resistance (secondary outcomes) in gestational week 37 among Swedish women.

Methods

Study Design

The HealthyMoms trial (clinicaltrials.gov NCT03298555) was a 2-arm parallel design randomized controlled trial conducted between October 2017 and November 2020 in the county of Östergötland, Sweden. The study received approval from the Regional Ethical Review Board in Linköping, Sweden (reference numbers 2017/112-31 and 2018/262-32) and all women provided written informed consent before entering the trial. Development of the HealthyMoms app and full details of the study design have been described previously [23]. The study is reported according to the Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and online Telehealth (CONSORT-EHEALTH) statement [24] (Multimedia Appendix 1).

Participants and Procedures

Between October 2017 and March 2020 participants were recruited in early pregnancy at the first routine visit at maternity clinics in the county of Östergötland, Sweden. During the study period approximately 4000 eligible women attended maternity care. At the maternity clinic, participants received written

information about the study, and women interested in participating contacted the research team via email or postal mail. Inclusion criteria were aged 18 years or older, a singleton pregnancy, and the ability to read and speak well-enough Swedish to be able to understand the app content. Women previously diagnosed with an eating disorder, diabetes, or other medical conditions with possible effects on body weight were excluded. Eligible women who agreed to participate were sent an accelerometer to assess physical activity and were instructed to register their diet using a web-based dietary assessment tool prior to the measurement. Baseline measures (13.9 [SD 0.7] gestational weeks) and follow-up measures (36.4 [SD 0.4] gestational weeks) were conducted at Linköping University Hospital. In short, these measures included assessment of body weight and height, body composition, plasma glucose, serum insulin, and sociodemographic variables. These are described in more detail below.

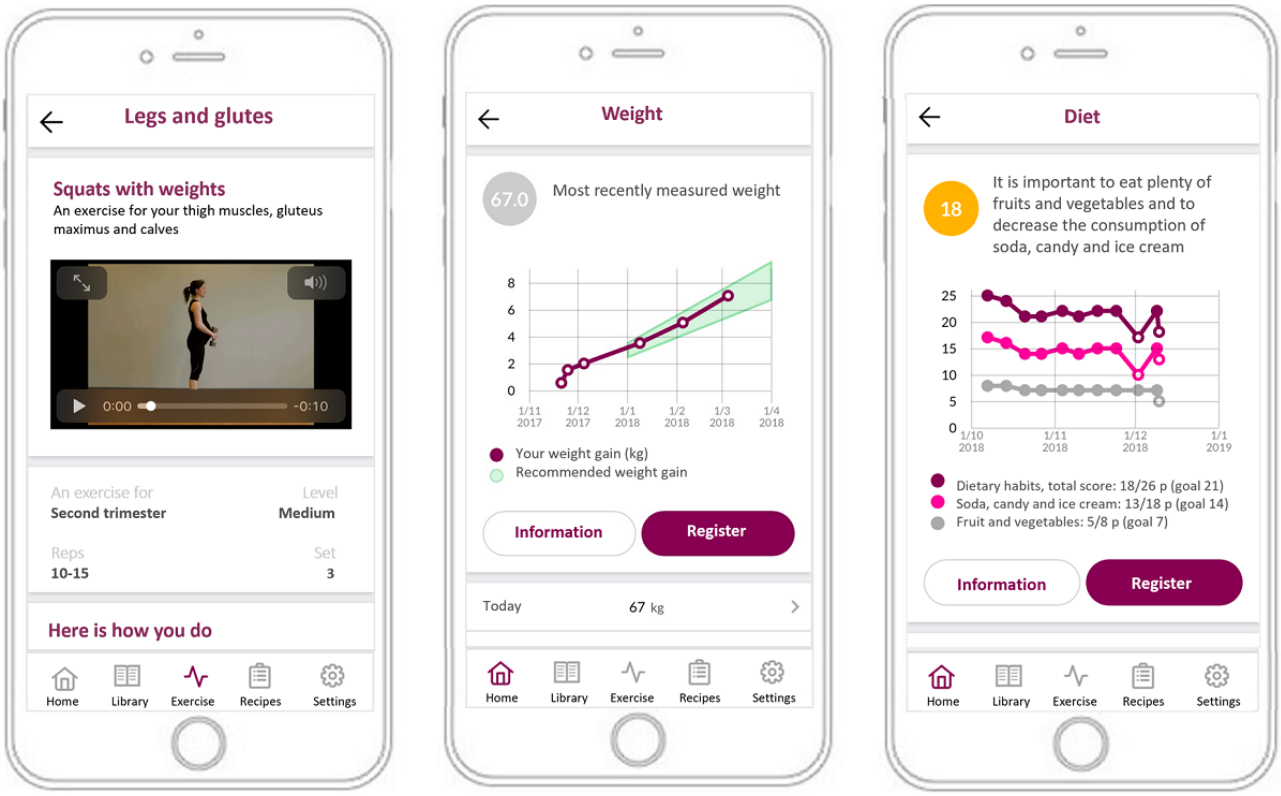
Control Group

The control group received standard maternity care consisting of regular monitoring of maternal and fetal health (such as measurements of blood pressure, blood glucose and ferritin, weight gain, symphysis fundus, as well as fetal movements and heart rate). Standard care also included an optional lecture in early pregnancy on a healthy lifestyle with some brief and general advice on diet, physical activity, smoking and alcohol, pregnancy-related health (eg, nausea, iron deficiency, pelvic pain), and medical care (eg, midwife visits, information on fetal diagnostics). In addition, standard care included repeated measurements of body weight throughout pregnancy.

Intervention

In addition to standard maternity care, participants in the intervention group received the HealthyMoms app (Android and iOS compatible), a 6-month program aimed at promoting recommended GWG [23] by encouraging a healthy diet and physical activity in accordance with current guidelines [25,26]. The app focuses on healthy dietary and physical activity habits as well as healthy GWG, irrespective of pre-pregnancy BMI. Figure 1 shows 3 screenshots from the HealthyMoms app. The development of the HealthyMoms app and its features has previously been described in detail [23]. In short, the app is grounded in social cognitive theory [27] and uses key behavior change techniques (eg, shaping knowledge, goal setting, feedback, and monitoring) which have been suggested to be important for promoting a healthy lifestyle also in pregnant women [15,28-30]. The app consists of 7 features, including informational themes that change every other week, push notifications, self-monitoring with feedback (for diet, physical activity, and GWG), recipes, exercise guide (eg, aerobic and resistance exercises and training programs) and videos, pregnancy calendar, and an app library (eg, frequently asked questions, practical tips). The themes address various topics, such as healthy food choices, exercise during pregnancy, a healthy GWG, and how to change habits. Participants also received automated push notifications 4 times/week with information, support, strategies, and guidance on how to achieve a behavior change and establish or maintain healthy habits (eg, improve diet and increase physical activity), as well as encouraging information, “take home messages” at the end of each theme, and reminders to use the self-monitoring features. The self-monitoring features provided the possibility to track weight gain, diet, and physical activity and to set a physical activity goal for MVPA (minutes per week). Participants received predesigned feedback based on their registrations of diet and physical activity in accordance with national guidelines [25,26]. Feedback consisted of a graphical visualization of the registrations where participants could review their reported diet, physical activity, and weight gain over time in relation to the recommendations. Participants also received feedback in text and a “traffic light” (ie, green: reached the recommendation; yellow: close to reaching the recommendation; red: far from reaching the recommendation) indicating compliance with recommendations following registration of diet and physical activity. Furthermore, the weight gain chart showed their individual GWG in relation to the recommended weight gain (according to the recommendations provided by the National Academy of Medicine [31], calculated from their self-reported pre-pregnancy BMI). Participants allocated to the intervention group were informed of the features in the app, how to download it (from AppStore [iOS] or Google Play [Android]), and were instructed to use the app as much as they preferred. Participants were registered as app users by the researchers, whereby they received an SMS text message with a link to a downloading site where they could download the app (free of charge). At follow-up, participants filled in a questionnaire regarding their satisfaction and usage of the app.

Figure 1. Screenshots from the HealthyMoms app showing examples from the app (ie, an exercise video, the weight gain chart, and diet registration with feedback).



Sample Size, Randomization, and Blinding

We estimated that 226 women (113 in each group) would provide 80% power (α=.05, 2 sided), assuming a common SD in GWG of 4 kg [32] to detect a difference of 1.5 kg between the control and intervention group. Considering a maximal loss to follow-up of 25% we recruited just over 300 women. After completion of baseline measures participants were randomized in a 1:1 ratio to either the control or intervention group using restricted randomization generated using STATA (version 13; StataCorp). Allocation concealment was ensured by using opaque envelopes (PH) which was opened by the assessor after completion of all the baseline measures whereby the participant was informed of their allocation (ie, the intervention or control group). The participants and outcome assessors were not blinded to the allocation due to the nature of the intervention.

Outcomes

The primary outcome was GWG between baseline (gestational week 14) and the follow-up measurement (gestational week 37). Secondary outcomes included body fatness, dietary habits (Swedish Healthy Eating Index), physical activity (time spent in MVPA), glycemia, and insulin resistance. All outcomes were assessed in gestational weeks 14 and 37.

Gestational Weight Gain

Body weight was measured after an overnight fast when the participant was wearing underwear using standardized procedures (Bod Pod; COSMED). Subsequently, GWG was calculated as the difference in body weight (in kg) between the baseline measurement (gestational week 14) and the follow-up measurement (gestational week 37). Furthermore, GWG between gestational weeks 14 and 37 was expressed per week (kg/week). Subsequently, we applied the pre-pregnancy BMI-specific GWG recommendations for the second and third trimester proposed by the National Academy of Medicine to categorize GWG for each woman as inadequate, adequate, and excessive (ie, underweight: 0.44-0.58 kg/week; normal weight: 0.35-0.50 kg/week; overweight: 0.23-0.33 kg/week; and obesity: 0.17-0.27 kg/week) [31].

Body Fatness

Body composition was measured using Bod Pod (COSMED) with accompanying software version 5.2.0 as described previously [33]. In short, the Bod Pod measures body volume using air-displacement plethysmography and body density was then derived by dividing body weight with body volume. Fat and fat-free mass were calculated using densities specified for the gestational age at baseline and follow-up measurements [34].

Dietary Habits

The web-based dietary recall method Riksmaten FLEX developed by the Swedish National Food Agency [35] adapted to pregnant women was used to assess dietary habits. The method utilized a repeated 24-hour recall approach over 3 days [35]. Upon the first log-in, participants were instructed to register their dietary intake for that day and the day before. The third day was automatically assigned to occur within 7 days of the first registration, on either a weekday or weekend day depending on what day the first day of registration was to ensure that registrations included both weekdays and weekend days [35].

The design of Riksmaten FLEX has been described in detail elsewhere [35]. In short, it consists of food items and prespecified dishes that participants can choose from. Once the correct food item or dish has been selected, participants defined portion sizes by choosing among pictures demonstrating various amounts of foods. Registrations were then linked to the Swedish National Food Composition Database, providing intakes of energy, macronutrients, and micronutrients [35]. In accordance with Moraeus et al [35], registered days with ≥3500 kcal or days ≤800 kcal were checked in detail by the research group to detect inaccurate energy intakes. In total, intakes for 1 (at baseline) and 3 days (at follow-up) were considered implausible and excluded. Intakes of selected food groups (ie, fruits, vegetables, red meat, fish, and shellfish) and macronutrients were summarized and averaged for each participant and day.

To assess diet quality, we calculated the Swedish Healthy Eating Index score [36], based on the Nordic Nutrition Recommendations [25], for each woman. The score consists of 9 components and was calculated based on intakes of fruit and vegetables (g/day), fish and shellfish (g/day), red meat (g/week), fiber (g/MJ), wholegrain (g/MJ), polyunsaturated fat (E%), monounsaturated fat (E%), saturated fat (E%), and sucrose (E%) as described elsewhere [36]. The score for each item ranged from 0 to 1 and the total score ranged from 0 to 9, with a higher score indicating better compliance with dietary guidelines [36].

Physical Activity

An ActiGraph wGT3x-BT (ActiGraph) accelerometer was used to assess physical activity. Participants were instructed to wear the accelerometer on the wrist for 7 consecutive 24-hour periods and to only remove it when engaging in water activities. The accelerometer was programmed to register accelerations at 100 Hz and participants filled in a diary where they reported sleep time and nonwear time which were used in the analysis to confirm sleep time. Participant data with at least one valid day were included. A valid day was defined as one-third or more of the 24-hour period being wear time, two-thirds or more of the wake time being wear time, and two-thirds or more of the sleep time being wear time. Participants who were not able to wear the accelerometer on the wrist (eg, health care workers due to hygiene restrictions at the workplace) were instructed to wear it on the hip instead, both at baseline and at follow-up (baseline n=23; follow-up n=18). Appropriate thresholds to identify MVPA were used for wrist- (ie, 100 m*g*) and hip-worn accelerometers (ie, 70 m*g*) [37]. Daily average MVPA was calculated as the weighted mean of weekdays and weekend days, that is, {([mean of weekdays × 5] + [mean of weekend days × 2])/7}. The intervention effect was very comparable when excluding the 29 women that wore the accelerometer on the hip or had less than 4 valid days [38] of recorded physical activity at both baseline and follow-up (results not shown). Data processing was conducted using the software program R [39] and the package GGIR [40].

Glycemia and Insulin Resistance

Blood samples were drawn after an overnight fast. Concentrations of glucose and insulin were analyzed on a Cobas 602 (Roche Diagnostics Scandinavia AB) at the Department of Clinical Chemistry, Linköping University Hospital. Insulin resistance was assessed by using the Homeostatic Model Assessment for Insulin Resistance (HOMA-IR) according to Matthews et al [41] and was calculated as fasting insulin [μU/L] × fasting glucose [mmol/L])/22.5.

Statistical Analysis

Effectiveness of the Intervention

All statistical analyses were conducted in accordance with the study protocol [23] and the CONSORT-EHEALTH statement [24]. Analyses followed principles of intention-to-treat and were performed in R version 3.6.3 (R Foundation for Statistical Computing). Null hypotheses were tested at the .05 significance level (2 sided). Missing data were imputed by means of multiple imputations in which the value of missing observations is predicted using available data. We used multiple imputations with chained equations [42] employing the analysis regression model as the prediction model within iterations. A total of 500 data sets were imputed for each analysis (predictive mean matching with 50 iterations) and analyses were pooled using Rubin's rules [43]. We also conducted complete case analyses for all outcomes. As described in the study protocol [23], we planned for per-protocol analysis including women who had used the app at least once and who had data on GWG at follow-up; however, only 1 participant was removed in the per-protocol analyses compared with the complete case analyses, and findings were unchanged. Therefore, we only report multiple imputations analyses and complete case analyses.

To contrast differences in primary (GWG in kg) and secondary outcomes (Swedish Healthy Eating Index, MVPA, body fatness, glycemia, and insulin resistance) between the 2 groups (intervention vs control) we estimated linear regression models. More specifically, for the primary outcome GWG, we regressed follow-up weight in gestational week 37 on group allocation and adjusted for baseline weight in gestational week 14 (crude model). This procedure has the advantage of being robust to imbalances at baseline and regression toward the mean [44]. The group coefficient in this model provides an estimate of the expected difference in GWG between 2 participants with the same baseline weight who have received different treatment (intervention vs control). Thereafter a second regression model was fitted with additional adjustments for pre-pregnancy BMI (underweight and normal weight vs overweight and obesity), parity (0 vs ≥1), and educational attainment (university degree vs no university degree) (adjusted model). Corresponding models were fitted for all secondary outcomes (Swedish Healthy Eating Index, physical activity, body fatness, glycemia, and insulin resistance). As planned, we also estimated effect modifications of the intervention on the primary outcome (GWG), by extending the regression model with interactions between group allocation and pre-pregnancy BMI, parity, and educational attainment, respectively.

Sensitivity and Complementary Analyses

We conducted the following sensitivity and complementary analyses. First, in accordance with our protocol [23], we conducted a sensitivity analysis excluding women diagnosed with gestational diabetes or pre-eclampsia (n=7) before the follow-up measurement and the intervention effect was

comparable (results not shown). Second, as in one of our recently published mHealth trials [45], we extended our prespecified statistical analysis plan with a Bayesian analysis for the primary outcome. Thus, we analyzed the interaction effect between group allocation and pre-pregnancy BMI using a Bayesian analysis [46]. This was done as the trial was not planned to be powered to detect this effect, and the Bayesian analysis allowed us to calculate the posterior probability of an interaction effect despite the null hypothesis not being rejected [47,48]. For the Bayesian analyses, imputation was done within the model estimation, that is, within the Markov Chain Monte Carlo simulations. For each iteration, missing data were replaced with draws from a normal distribution specified by the sample parameters of the regression model being estimated.

Results

Participants

As shown in Figure 2, approximately 4000 eligible pregnant women entered maternity health care during the recruitment period (October 2017 to March 2020), and 399 women expressed interest to the research group to participate in the study. Ninety-four women were excluded because (1) they did not meet all the inclusion criteria (n=21), (2) declined participation after full information (n=27), (3) had a miscarriage before enrollment (n=25), or (4) did not reply after expressing interest to participate (n=21). A total of 305 pregnant women were enrolled and randomized and their baseline characteristics are provided in Table 1. At baseline, their mean age was 31 (SD 4) years; 57.4% (175/305) were nulliparous; and 2.0% (6/305) were underweight, 69.5% (212/305) normal weight, and 28.5% (87/305) had overweight or obesity pre-pregnancy. As shown in Table 1, there were no major differences in baseline characteristics between the women in the intervention and control group. In total, 152 participants were allocated to the intervention group, of which 151 downloaded the app, and all participants were included in the analyses regardless of app usage. Table 2 reports the self-reported satisfaction of the HealthyMoms app. For instance, 77.6% (104/134) fully or largely agreed with the statement that they were satisfied with the app and another 13.4% (18/134) agreed to some extent. Additionally, the majority of participants (82.8%, 111/134) reported using the app at least once per week and only a few (17.2%, 23/134) reported using the app 2 to 3 times per month or less. Additional subjective data on app usage are provided in Multimedia Appendix 2. Furthermore, objective measures showed the following mean usage of the registration features: physical activity: 1.6 (SD 2.1) times/week; self-monitoring for weight: 0.7 (SD 0.8) times/week; and diet registration 0.2 (SD 0.3) times/week.

Figure 2. Flowchart of the HealthyMoms trial.

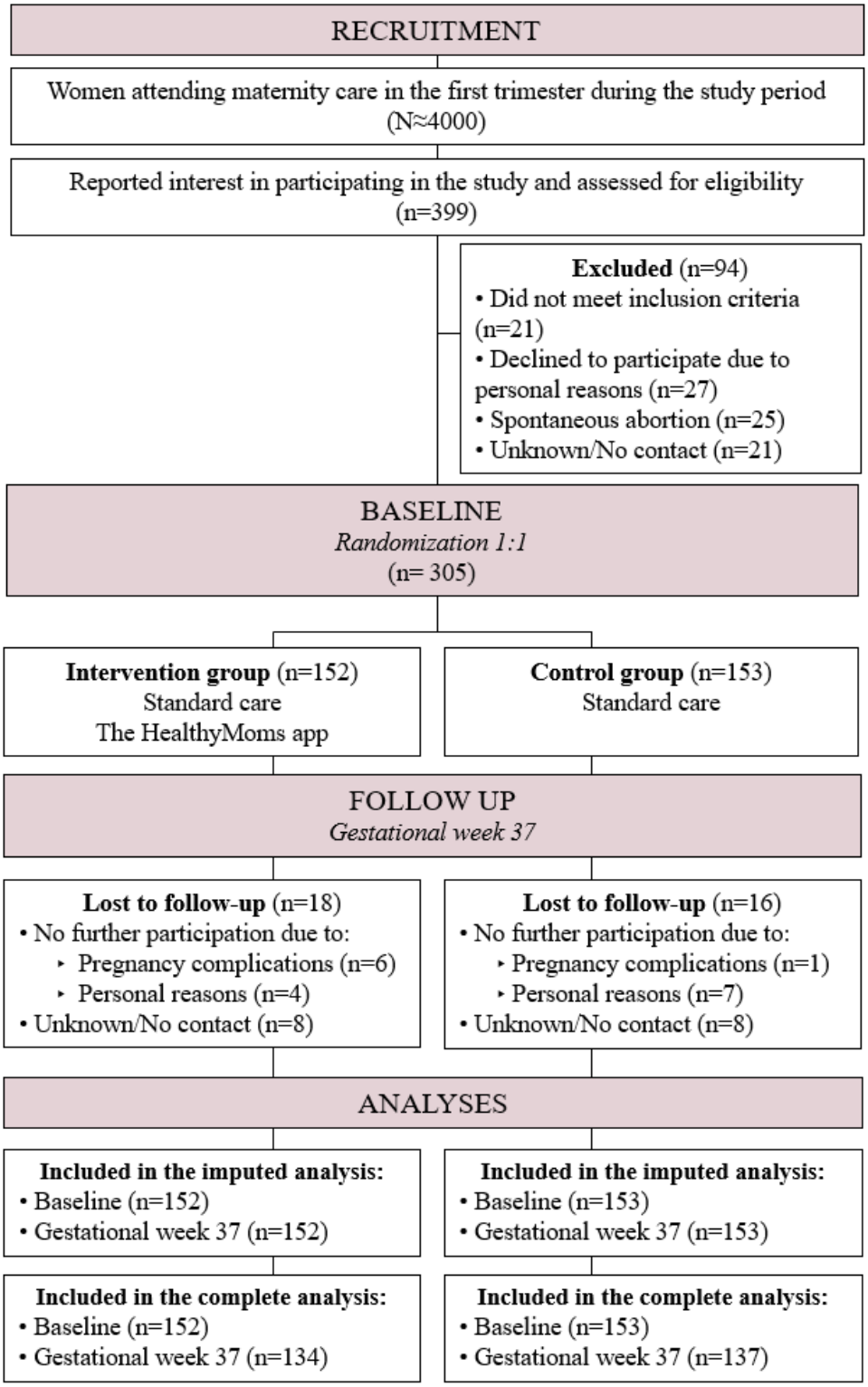


Table 1. Baseline characteristics of the women in the HealthyMoms trial.

Characteristics	All women (n=305)	Intervention (n=152)	Control (n=153)
Pre-pregnancy characteristics, n (%)			
Parity[a]			
0	175 (57.4)	86 (56.6)	89 (58.2)
≥1	130 (42.6)	66 (43.4)	64 (41.8)
Educational attainment[a]			
Primary school (9 years)	2 (0.7)	0 (0.0)	2 (1.3)
High school (12 years)	66 (21.6)	37 (24.3)	29 (19.0)
University degree	237 (77.7)	115 (75.7)	122 (79.7)
Pre-pregnancy BMI[a,b]			
Underweight (<18.5 kg/m^2)	6 (2.0)	1 (0.7)	5 (3.3)
Normal weight (18.5-24.9 kg/m^2)	212 (69.5)	103 (67.8)	109 (71.2)
Overweight (25.0-29.9 kg/m^2)	67 (22.0)	34 (22.4)	33 (21.6)
Obesity (≥30.0 kg/m^2)	20 (6.6)	14 (9.2)	6 (3.9)
Measured variables in gestational week 14 (baseline), mean (SD)			
General characteristics[a]			
Gestational week	13.9 (0.7)	13.8 (0.6)	14.0 (0.7)
Age (years)	31.3 (4.1)	31.4 (4.3)	31.3 (3.8)
Anthropometry[a]			
Weight (kg)	67.7 (11.5)	68.3 (12.8)	67.0 (10.2)
Height (m)	1.67 (0.06)	1.66 (0.06)	1.68 (0.06)
BMI (kg/m^2)	24.2 (3.8)	24.7 (4.3)	23.8 (3.2)
Fat mass index (kg/m^2)	8.0 (3.2)	8.4 (3.6)	7.6 (2.6)
Fat free mass index (kg/m^2)	16.2 (1.3)	16.2 (1.4)	16.2 (1.3)
Swedish Healthy Eating Index Score[c]	6.66 (0.98)	6.54 (0.98)	6.79 (0.97)
Moderate-to-vigorous physical activity (min/day)[d]	39.2 (24.0)	38.7 (24.6)	39.8 (23.5)
Glycemia (mmol/L)[e]	4.8 (0.3)	4.8 (0.3)	4.8 (0.3)
Homeostatic Model Assessment for Insulin Resistance[e]	1.4 (0.7)	1.4 (0.8)	1.4 (0.7)

[a]All women (n=305): intervention group (n=152) + control group (n=153).

[b]Based on self-reported pre-pregnancy weight and height.

[c]All women (n=302): intervention group (n=151) + control group (n=151).

[d]All women (n=296); intervention group (n=146) + control group (n=150).

[e]All women (n=304); intervention group (n=151) + control group (n=153).

Table 2. Self-reported app satisfaction in the intervention group (n=134) at the follow-up measurement. Participants responded to the following statements with the 6 alternatives shown.

Statement	Strongly disagree, n (%)	Agree to a small extent, n (%)	Agree to some extent, n (%)	Strongly agree, n (%)	Fully agree, n (%)	Do not know, n (%)
I am satisfied with the app	2 (1.5)	5 (3.7)	18 (13.4)	66 (49.3)	38 (28.4)	5 (3.7)
The app has been a good support for a healthy weight gain during pregnancy	9 (6.7)	18 (13.4)	39 (29.1)	32 (23.9)	20 (14.9)	16 (11.9)
The app has been a good support for healthy food habits	12 (9.0)	16 (11.9)	42 (31.3)	41 (30.6)	11 (8.2)	12 (9.0)
The app has been a good support for exercise habits	15 (11.2)	16 (11.9)	29 (21.6)	44 (32.8)	20 (14.9)	10 (7.5)
The app has given me insight regarding my food habits	26 (19.4)	16 (11.9)	39 (29.1)	31 (23.1)	9 (6.7)	13 (9.7)
The app has given me insight regarding how physically active I am	28 (20.9)	16 (11.9)	32 (23.9)	36 (26.9)	13 (9.7)	9 (6.7)
I think that the HealthyMoms app is better than other similar apps	3 (2.2)	10 (7.5)	31 (23.1)	24 (17.9)	9 (6.7)	57 (42.5)
I would recommend other pregnant women to use the HealthyMoms app	3 (2.2)	7 (5.2)	16 (11.9)	45 (33.6)	57 (42.5)	6 (4.5)

Effectiveness of the Intervention (Primary Outcomes)

Table 3 presents the intervention effects for the primary outcome (GWG) for the crude model as well as the model adjusted for pre-pregnancy BMI, parity, and educational attainment for the imputed (n=305) and the complete cases analyses (n=271), respectively. Overall, results showed no statistically significant difference between the groups on GWG (–0.20 kg; 95% CI –0.98 to 0.59; *P*=.62 for the intervention group vs the control, n=305). Furthermore, as shown in Table 3, results were similar when taking baseline weight, pre-pregnancy BMI, parity, and educational attainment into account as well as in analyses with complete cases. Overall, 13.3% (36/271), 36.9% (100/271), and 49.8% (135/271) of the women gained below, within, and above the recommendations by the National Academy of Medicine, and there was no statistical difference in adherence to the recommendations between the intervention and control group (Table 4).

Table 3. Intervention effect on gestational weight gain (primary outcome) assessed using regression analysis[a,b,c].

Model	Imputed data analysis (n=305)		Complete cases analysis (n=271)	
	Coefficient (95% CI)	*P* value	Coefficient (95% CI)	*P* value
Crude	–0.20 (–0.98 to 0.59)	.62	–0.22 (–1.00 to 0.56)	.58
Adjusted	–0.20 (–1.00 to 0.60)	.62	–0.24 (–1.01 to 0.54)	.55

[a]Regression analysis of follow-up measure of weight on group allocation. The coefficient is interpreted as the estimated effect of the intervention compared with the control adjusted for baseline weight (crude model), BMI category (underweight and normal weight vs overweight and obesity), parity (0 vs 1 or more), and educational attainment (university degree vs no university degree) (adjusted model).

[b]Baseline, n=305 (152 intervention and 153 control); Follow-up, n=271 (134 intervention and 137 control).

[c]At baseline, the mean bodyweight (kg) for the intervention and control group was 68.3 (SD 12.8) and 67.0 (SD 10.2), respectively, whereas at follow-up the corresponding values were 78.7 (SD 13.1) and 77.3 (SD 10.6).

Table 4. Intervention effect on gestational weight gain according to National Academy of Medicine's recommendations.

Outcome	Descriptive data, n (%)		Intervention effect using regression analysis[a]			
	Group		Imputed data analysis (n=305)		Complete cases analysis (n=271)	
	Intervention (n=134)	Control (n=137)	Odds ratio[a] (95% CI)	*P* value	Odds ratio[a] (95% CI)	*P* value
Excessive GWG[b,c]	67 (50.0)	68 (49.6)	0.75 (0.43-1.32)	.31	0.75 (0.43-1.32)	.32
Adequate GWG[b]	52 (38.8)	48 (35.0)	Reference		Reference	
Inadequate GWG[b]	15 (11.2)	21 (15.3)	0.66 (0.30-1.43)	.29	0.66 (0.30-1.44)	.29

[a]Regression analysis of gestational weight gain on group allocation. The coefficient is interpreted as the estimated effect of the intervention compared with the control adjusted for baseline body weight, BMI category (underweight and normal weight vs overweight and obesity), parity (0 vs 1 or more), and educational attainment (university degree vs no university degree).

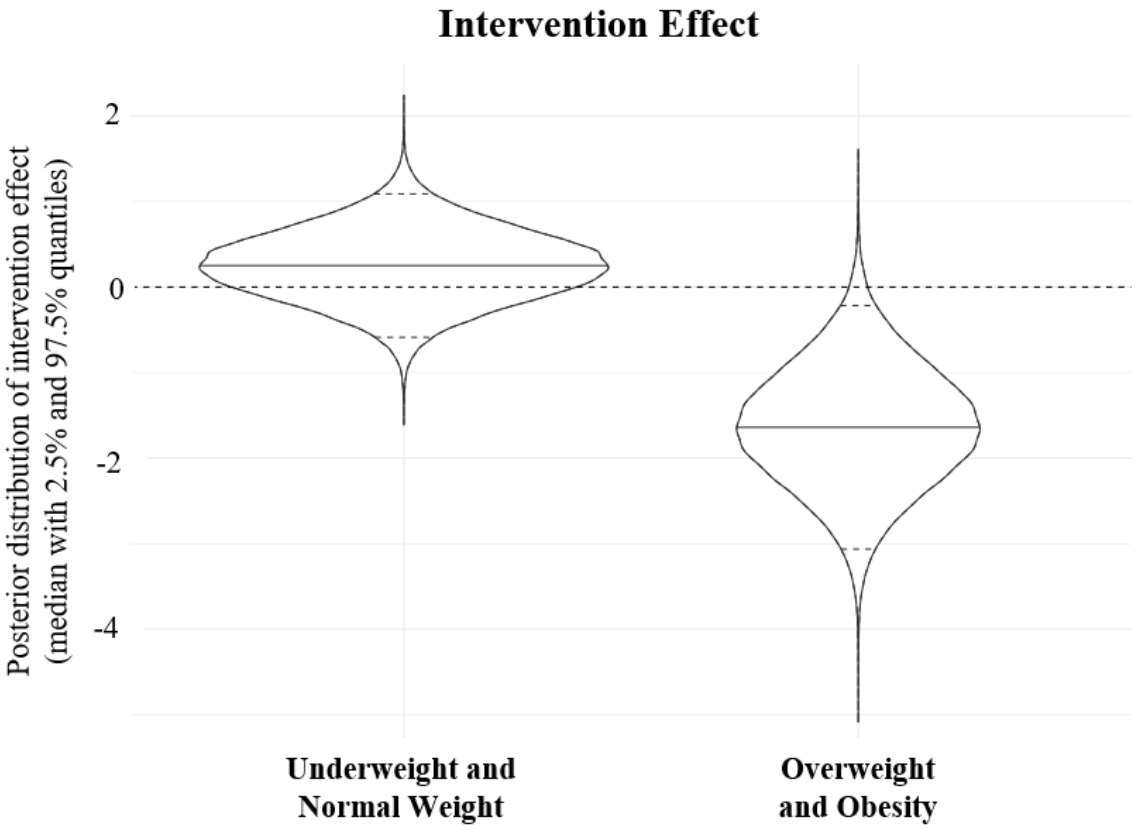
[b]GWG was calculated as the difference between weight at follow-up and baseline, which then was divided by gestational weeks to obtain GWG expressed as kg/week. This GWG (kg/week) was compared to the weekly GWG recommendations by the National Academy of Medicine for the second and third trimesters to classify GWG as excessive, adequate, or inadequate.

[c]GWG: gestational weight gain.

There was no statistically significant interaction effect for parity or educational attainment (results not shown); however, data indicated that there was a marked interaction between pre-pregnancy BMI and group allocation with the intervention being more effective in women with overweight and obesity compared with those who were underweight and normal weight. Thus, for women with overweight and obesity, GWG in the intervention group was 1.33 kg (95% CI –2.92 to 0.26, *P*=.10, n=305) lower than those in the control group when also accounting for parity and educational attainment. In the complete case analysis, the interaction effect was stronger and statistically significant (–1.67 kg; 95% CI –3.26 to –0.09; *P*=.031, n=271).

The interaction effect was furthermore supported by the results from a Bayesian estimation of the same interaction model (Figure 3). The probability that the expected GWG in the intervention group was less than that in the control group (ie, the intervention had any effect on GWG) was only 27% among underweight and normal weight women; however, among women with overweight and obesity it was 99%. Furthermore, the probability that this effect was over 1 and 1.5 kg among women with overweight and obesity was 81% and 57%, respectively.

Figure 3. Bayesian analysis (with imputation, n=305) of the intervention effect on gestational weight gain according to prepregnancy BMI.



Prepregnancy BMI category	Effect estimate[a]	Posterior probability of estimate[a] being lower than 0 kg, -1 kg, -1.5 kg, and -2 kg			
	Median (kg) (2.5%; 97.5%)	< 0 kg	< -1 kg	< -1.5 kg	< -2 kg
Underweight or normal weight before pregnancy	0.25 (-0.58; 1.09)	27%	< 0.01%	< 0.01%	< 0.01%
Overweight or obesity before pregnancy	-1.63 (-3.05; -0.21)	99%	81%	57%	31%

[a]The regression model included follow-up weight regressed on group allocation, prepregnancy BMI category, and the interaction of group allocation × BMI category (underweight and normal weight vs overweight and obesity), baseline weight, parity (0 vs 1 or more), and educational attainment (university degree vs no university degree).

Secondary Outcomes

Table 5 shows the corresponding results for the secondary outcomes. The intervention group had higher scores for the Swedish Healthy Eating Index at follow-up than the control group (0.27; 95% CI 0.05-0.50; *P*=.017) when adjusting for baseline values as well as for pre-pregnancy BMI, parity, and educational attainment. The difference in the Swedish Healthy Eating Index was driven by an overall shift, with slightly higher scores in 7 out of 9 components, toward a healthier diet as indicated by higher intakes of fruits and vegetables and with a statistically significant observed reduction in the intake of red meat (*P*=.027) (Multimedia Appendix 3). No statistically significant differences in MVPA or any of the other secondary outcomes at gestational week 37 were observed between the intervention and control group (*P*≥.21).

Table 5. Intervention effect on the secondary outcomes.

Outcome	Descriptive data, mean (SD)		Intervention effect using regression analysis[a]			
	Group		Imputed data analysis		Complete cases analysis	
	Intervention	Control	Coefficient[a] (95% CI)	*P* value	Coefficient[a] (95% CI)	*P* value
Swedish Healthy Eating index (points)						
Baseline (n=302)[b]	6.54 (0.98)	6.79 (0.97)	0.27 (0.05 to 0.50)	.017	0.27 (0.05 to 0.50)	.018
Follow-up (n=269)[b]	6.53 (0.94)	6.38 (1.07)				
Moderate-to-vigorous physical activity (min/day)						
Baseline (n=296)[c]	38.7 (24.6)	39.8 (23.5)	–0.76 (–5.34 to 3.80)	.74	–1.01 (–5.66 to 3.62)	.67
Follow-up (n=267)[c]	26.3 (19.0)	27.8 (24.7)				
Fat mass (kg)						
Baseline (n=305)[d]	23.4 (10.1)	21.5 (7.6)	0.05 (–0.65 to 0.76)	.88	–0.03 (–0.71 to 0.64)	.92
Follow-up (n=268)[d]	26.8 (9.7)	24.7 (7.5)				
Fat-free mass (kg)						
Baseline (n=305)[d]	45.0 (4.8)	45.6 (4.8)	–0.09 (–0.46 to 0.28)	.64	–0.07 (–0.45 to 0.30)	.70
Follow-up (n=268)[d]	51.9 (5.4)	52.5 (5.4)				
Glycemia (mmol/L)						
Baseline (n=304)[e]	4.8 (0.3)	4.8 (0.3)	0.06 (–0.03 to 0.15)	.21	0.06 (–0.03 to 0.14)	.18
Follow-up (n=263)[e]	4.7 (0.5)	4.6 (0.3)				
Homeostatic Model Assessment for Insulin Resistance						
Baseline (n=304)[e]	1.41 (0.76)	1.36 (0.65)	0.10 (–0.13 to 0.34)	.39	0.12 (–0.11 to 0.36)	.31
Follow-up (n=263)[e]	2.41 (1.36)	2.19 (0.98)				

[a]Regression analysis of follow-up measure of secondary outcome on group allocation. The coefficient is interpreted as the estimated effect of the intervention compared with the control adjusted for baseline value of the secondary outcome, BMI category (underweight and normal weight vs overweight and obesity), parity (0 vs 1 or more), and educational attainment (university degree vs no university degree). Imputed data analysis included data for all 305 women and the complete cases analysis included data for 263-269 women.

[b]Baseline, n=302 (151 intervention and 151 control); follow-up, n=269 (135 intervention and 134 control).

[c]Baseline, n=296 (146 intervention and 150 control); follow-up, n=267 (132 intervention and 135 control). Number of valid days for accelerometry: intervention group (baseline: 6.5 [SD 1.1] days; follow-up: 6.7 [SD 0.8] days); control group (baseline: 6.7 [SD 0.9] days; follow-up: 6.7 [SD 1.1] days). Average wear time for valid days: intervention group (baseline: 99.0%; follow-up: 97.5%); control group (baseline: 98.7%; follow-up: 98.4%).

[d]Baseline, n=305 (152 intervention and 153 control); follow-up, n=268 (133 intervention and 135 control).

[e]Baseline, n=304 (151 intervention and 153 control); follow-up, n=263 (130 intervention and 133 control).

Discussion

Principal Findings

This study is the first to examine the effectiveness of a comprehensive intervention delivered solely via an app on GWG, body fatness, dietary habits, physical activity, glycemia, and insulin resistance in pregnant women from all BMI categories. We did not observe any statistically significant effect on GWG; however, there was some evidence that women with overweight and obesity before pregnancy gained less weight in the intervention group as compared with the control group in the imputed analyses (–1.33 kg; 95% CI –2.92 to 0.26; *P*=.10) and the completers-only analyses (–1.67 kg; 95% CI –3.26 to –0.09; *P*=.031). Furthermore, a Bayesian analysis supported that the intervention was more effective among women with overweight and obesity. Regarding secondary outcomes, we found no effect on body fatness, MVPA, glycemia, and insulin resistance; however, participants in the intervention group had a higher score in the Swedish Healthy Eating Index in gestational week 37 compared with the control group.

Comparison With Previous Studies

Previous studies of apps promoting healthy GWG have been pilot studies [15] and have only included women with overweight and obesity. Nevertheless, our study can be compared with the 3-arm randomized controlled trial by Olson et al [19] in pregnant women in the United States (n=1689, BMI ≥18.5-35 kg/m^2) evaluating the effectiveness of access to a website aimed to support healthy weight gain, diet, and physical activity in pregnancy. Similar to Olson et al [19], we found no statistically significant effect on total GWG and the participants

were primarily normal weight preconception. Interestingly, as stated above, we observed an interaction between pre-pregnancy BMI and group allocation on GWG, suggesting that women with overweight and obesity benefitted more from the intervention. These results were further complemented by our Bayesian analysis showing a 99% probability that the intervention had any effect and an 81% probability that this effect was larger than 1 kg. Our results are also in line with a previous pilot study of an mHealth intervention among pregnant women with overweight and obesity [18], where participants in the intervention group gained less weight than those in the control group (7.8 kg [SD 4.7] kg vs 9.7 kg [SD 3.9]; *P*=.041). Additionally, the intervention effect estimated in our study using only an app is comparable to previous interventions in pregnant women with overweight and obesity relying on traditional face-to-face counseling [10]. This suggests that mHealth interventions may be as effective as these traditional approaches, while using less resources, being more cost-effective, and having greater reachability. Taken together, mHealth interventions have potential to be made readily available to many women at the touch of a button.

Another main finding is that we observed a statistically significant higher Swedish Healthy Eating Index score in the intervention group compared with the control group. As shown in Multimedia Appendix 3, from the individual components of the score, the higher score was driven by an overall shift toward a healthier diet, including a statistically significant reduction in the intake of red meat. The effect on red meat may be explained by the fact that the HealthyMoms app was carefully designed to also highlight the benefits of plant-based diets by providing information and weekly menus with only vegetarian foods (in addition to the mixed diet menus) in the recipe module. Comparisons with previous mHealth studies aiming to improve dietary behaviors are difficult because dietary interventions are complex with various focus and intervention features. Additionally, studies have used different outcome measures (eg, food frequency questionnaires that may be less sensitive to capture variations in dietary intake [18]). Nevertheless, it is interesting to note that similar to our study, Ainscough et al [22] as well as Li et al [49] observed statistically significant effects on dietary outcomes using digital interventions in pregnant women with overweight and obesity. In summary, our results provide evidence that a 6-month behavioral intervention through an app could potentially improve dietary behaviors in pregnant women and thus represent a key target for future mHealth GWG interventions.

In contrast to the positive findings for dietary behaviors, we observed no effect on MVPA in this study. A few pilot mHealth studies in pregnant women have shown a beneficial effect on physical activity [18,50], while others have not [51]; however, these studies evaluated self-reported outcomes [18] or utilized steps from Fitbit manually imputed by participants [51] or had a short duration [50]. Noteworthy, our results showing an effect on diet but not on MVPA may be reconciled with previous face-to-face interventions which have shown stronger effects on GWG for interventions targeting dietary behaviors compared with physical activity [52]. It is possible that the relatively modest effect on diet coupled with the lack of an effect on physical activity may explain why the intervention did not have any overall effect on GWG. One potential explanation for not observing an effect on MVPA could be that the third trimester is characterized by an increase in body mass and fetal growth which may inhibit the capacity to perform certain exercises, as well as symptoms such as pelvic pain that may decrease the ability or the motivation to be physically active. Thus, at this stage of gestation, reductions in physical activity are common [53] and it may be difficult to detect if the intervention has counteracted a natural decline. In this context it is also relevant to note that Choi et al [51] reported that it was difficult to recruit inactive women that wanted to increase physical activity during pregnancy for an intervention. Consequently, further studies to explore the potential of mHealth interventions to improve physical activity behaviors during pregnancy are required and preferably such studies should include objective measures of physical activity at multiple timepoints from early pregnancy.

Strengths and Limitations

The HealthyMoms trial has several strengths. These include the randomized control design, high compliance (88.8% completion rate, 271/305) which provided adequate power to assess our outcomes, and the use of accurate and objective methods to assess primary and secondary outcomes (ie, measurements of body weight and body fatness using Bod Pod and objective measures of physical activity). Furthermore, another strength is that the intervention is theory-informed and uses key behavioral change techniques [27,28] which have been shown to be important elements with regard to intervention effectiveness and engagement in pregnant women [15]. More specifically, we utilized goal setting, self-monitoring, and feedback which have been demonstrated to increase effectiveness of digital dietary and physical activity interventions targeting pregnant women [15]. An additional strength of the study is the inclusion of women from all BMI categories and not only women with overweight and obesity. This study has also limitations to be acknowledged. Our study sample contained a larger proportion of women with a university degree compared with women in the general population (78% vs approximately 47%) [54]. This may influence generalizability to women with lower socioeconomic status although we found no evidence that the intervention effect differed according to educational attainment. Correspondingly, the prevalence of women with overweight and obesity in this study was somewhat lower than the general pregnant population in Sweden (29% vs 37%) [4]. Nevertheless, the prevalence of inadequate and excessive GWG in the trial (13% inadequate, 50% excessive) was similar to the general population (18% inadequate, 47% excessive) [4]. Besides, because we did not conduct randomization stratified by pre-pregnancy BMI, the number of women with overweight/obesity was slightly different in the intervention and control group which may somewhat decrease statistical power. Nevertheless, we were able to observe effects on GWG in women with overweight and obesity, although future studies could include a larger sample of such women. Furthermore, an inherent limitation is the risk of recall bias and social desirability in dietary assessments since established methods rely on self-report. However, we utilized a comprehensive and feasible web-based 24-hour recall method, which has previously been

used in a national survey in Sweden [35]. Compared with other commonly used methods, 24-hour recalls are to a lower extent associated with misreporting of dietary intakes [55]. Although we mainly utilized objective study measures, another limitation is that the assessors were only masked to the randomization assignment for the baseline measures. Furthermore, it cannot be excluded that the intervention effect was diluted as the control group was carefully measured in terms of body composition and weight, diet, and physical activity, in order to compare the groups regarding outcome measures, which may have influenced their behavior. Finally, another limitation of the study is that the intervention is only available in Swedish and thus women not proficient in the Swedish language were excluded. Thus, topics for future research should include pregnant women from various socioeconomic and migrant backgrounds [56].

Clinical and Public Health Relevance

The HealthyMoms trial provides several important findings. Although we did not observe a statistically significant overall effect on the primary outcome (GWG), our findings indicate a meaningful effect in pregnant women with overweight or obesity compared with standard care, with a similar effect as seen in traditional face-to-face interventions [10]. Furthermore, we observed a positive overall effect of the intervention on dietary habits, regardless of pre-pregnancy BMI. Self-reported data on app usage indicated high engagement and satisfaction by participants, which was also confirmed in a detailed qualitative [57] and a quantitative evaluation in a subsample (to be reported elsewhere). Altogether, these results indicate that the HealthyMoms app has potential to be a valuable tool to promote healthy diet and weight gain in pregnant women. Access to scalable and cost-effective interventions such as the HealthyMoms app may be of importance in order to counteract the high prevalence of excessive GWG in Sweden and globally [1,2,4] and its negative implications on pregnancy complications and long-term maternal and fetal health [2,5]. More specifically, the HealthyMoms app is particularly attractive in this context because it is a fully automated stand-alone intervention, which does not rely on additional resources from health care. The next step is to estimate the effectiveness of the HealthyMoms app across Sweden including women from various socioeconomic and migrant backgrounds as well as additional pregnancy outcomes such as pre-eclampsia and gestational diabetes. If clinical variables collected routinely in health care are used as outcomes, access to the app could be given already in the first trimester which may increase the effectiveness of the HealthyMoms app. Altogether, such an evaluation will be important considering the great lack of large-scale digital interventions in this field.

Conclusions

Although we found no overall effect on GWG, our results demonstrate the potential of a smartphone app (HealthyMoms) to promote healthy dietary behaviors as well as to decrease weight gain during pregnancy in women with overweight and obesity, and with a similar effect as in traditional interventions. Thus, this intervention, solely delivered through an app, has potential to be useful for promoting a healthy lifestyle during pregnancy in many women while using less resources from health care.

Acknowledgments

This study was funded by the Swedish Research Council (2016-01147 to ML) and additionally supported by the Swedish Research Council for Health, Working Life and Welfare (Forte, 2017-00088 to PH; 2018-01410 to ML); Bo and Vera Ax:son Johnsons' Foundation (to ML); the Strategic Research Area Health Care Science, Karolinska Institutet/Umeå University (to PH); the Swedish Society of Medicine (to PH); Karolinska Institutet (to PH); Lions Forskningsfond (to PH); and ALF Grants, Region Östergötland (ML). MHL was supported by a grant from Yrjö Jahnsson Foundation. The funders had no role in study design, conduct, or reporting of the trial. The authors thank all the participants in the trial as well as Eva Flinke and Ellinor Nilsson for invaluable help with data collection, the midwives in Region Östergötland for help with the recruitment, Hanna Henriksson and Christina Alexandrou for valuable help developing the content of the app, Jan Fjellström and Nils Lidström at ScientificMed for development of the app and technical support, and Anna Karin Lindroos and Eva Warensjö Lemming for support with the Riksmaten FLEX method (adaptations, data processing, and interpretation of results).

Authors' Contributions

ML is the principal investigator of this randomized controlled trial. The study was conceptualized and designed together with PH, MBL, and RM. JS was responsible for developing the intervention content with support from PH and ML. JS and ES were responsible for the recruitment and data collection. JHM and MH contributed to processing of accelerometer and dietary data, respectively. MB contributed to data analysis with statistical expertise. ML, JS, PH, MBL, and MHL contributed to the development of the HealthyMoms app. JS, ES, PH, and ML drafted the manuscript and all authors have approved the final draft and submitted manuscript.

Conflicts of Interest

The authors declare no conflict of interest. MB owns a private company (Alexit AB) that works with development and dissemination of eHealth apps to health organizations and professionals (both private and public sector). Alexit AB was not involved in this study.

Multimedia Appendix 1
CONSORT eHealth checklist.
[PDF File (Adobe PDF File), 706 KB - mhealth_v9i3e26091_app1.pdf]

Multimedia Appendix 2
Additional self-reported data on app usage and satisfaction in the intervention group (n=134) at the follow-up measurement. Participants responded to the following statements with the 6 alternatives shown.
[PDF File (Adobe PDF File), 27 KB - mhealth_v9i3e26091_app2.pdf]

Multimedia Appendix 3
Intervention effect on the components in the Swedish Healthy Eating Index.
[PDF File (Adobe PDF File), 28 KB - mhealth_v9i3e26091_app3.pdf]

References

1. Goldstein RF, Abell SK, Ranasinha S, Misso ML, Boyle JA, Harrison CL, et al. Gestational weight gain across continents and ethnicity: systematic review and meta-analysis of maternal and infant outcomes in more than one million women. BMC Med 2018 Aug 31;16(1):153 [FREE Full text] [doi: 10.1186/s12916-018-1128-1] [Medline: 30165842]
2. Goldstein RF, Abell SK, Ranasinha S, Misso M, Boyle JA, Black MH, et al. Association of Gestational Weight Gain With Maternal and Infant Outcomes: A Systematic Review and Meta-analysis. JAMA 2017 Jun 06;317(21):2207-2225 [FREE Full text] [doi: 10.1001/jama.2017.3635] [Medline: 28586887]
3. Henriksson P, Eriksson B, Forsum E, Löf M. Gestational weight gain according to Institute of Medicine recommendations in relation to infant size and body composition. Pediatr Obes 2015 Oct;10(5):388-394. [doi: 10.1111/ijpo.276] [Medline: 25521831]
4. Henriksson P, Sandborg J, Blomberg M, Nowicka P, Petersson K, Bendtsen M, et al. Body mass index and gestational weight gain in migrant women by birth regions compared with Swedish-born women: A registry linkage study of 0.5 million pregnancies. PLoS One 2020 Oct 29;15(10):e0241319 [FREE Full text] [doi: 10.1371/journal.pone.0241319] [Medline: 33119672]
5. LifeCycle Project-Maternal ObesityChildhood Outcomes Study Group, Voerman E, Santos S, Inskip H, Amiano P, Barros H, et al. Association of Gestational Weight Gain With Adverse Maternal and Infant Outcomes. JAMA 2019 May 07;321(17):1702-1715 [FREE Full text] [doi: 10.1001/jama.2019.3820] [Medline: 31063572]
6. Han Z, Lutsiv O, Mulla S, Rosen A, Beyene J, McDonald SD, Knowledge Synthesis Group. Low gestational weight gain and the risk of preterm birth and low birthweight: a systematic review and meta-analyses. Acta Obstet Gynecol Scand 2011 Sep;90(9):935-954 [FREE Full text] [doi: 10.1111/j.1600-0412.2011.01185.x] [Medline: 21623738]
7. International Weight Management in Pregnancy (i-WIP) Collaborative Group. Effect of diet and physical activity based interventions in pregnancy on gestational weight gain and pregnancy outcomes: meta-analysis of individual participant data from randomised trials. BMJ 2017 Jul 19;358:j3119 [FREE Full text] [doi: 10.1136/bmj.j3119] [Medline: 28724518]
8. Peaceman AM, Clifton RG, Phelan S, Gallagher D, Evans M, Redman LM, LIFE‐Moms Research Group. Lifestyle Interventions Limit Gestational Weight Gain in Women with Overweight or Obesity: LIFE-Moms Prospective Meta-Analysis. Obesity (Silver Spring) 2018 Sep;26(9):1396-1404 [FREE Full text] [doi: 10.1002/oby.22250] [Medline: 30230252]
9. Muktabhant B, Lawrie TA, Lumbiganon P, Laopaiboon M. Diet or exercise, or both, for preventing excessive weight gain in pregnancy. Cochrane Database Syst Rev 2015 Jun 15(6):CD007145. [doi: 10.1002/14651858.CD007145.pub3] [Medline: 26068707]
10. Shieh C, Cullen DL, Pike C, Pressler SJ. Intervention strategies for preventing excessive gestational weight gain: systematic review and meta-analysis. Obes Rev 2018 Aug;19(8):1093-1109. [doi: 10.1111/obr.12691] [Medline: 29806187]
11. Lau Y, Klainin-Yobas P, Htun TP, Wong SN, Tan KL, Ho-Lim ST, et al. Electronic-based lifestyle interventions in overweight or obese perinatal women: a systematic review and meta-analysis. Obes Rev 2017 Sep;18(9):1071-1087. [doi: 10.1111/obr.12557] [Medline: 28544551]
12. Mateo GF, 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]
13. Schoeppe S, Alley S, Lippevelde WV, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
14. Redman LM, Gilmore LA, Breaux J, Thomas DM, Elkind-Hirsch K, Stewart T, et al. Effectiveness of SmartMoms, a Novel eHealth Intervention for Management of Gestational Weight Gain: Randomized Controlled Pilot Trial. JMIR Mhealth Uhealth 2017 Sep 13;5(9):e133 [FREE Full text] [doi: 10.2196/mhealth.8228] [Medline: 28903892]

15. Rhodes A, Smith AD, Chadwick P, Croker H, Llewellyn CH. Exclusively Digital Health Interventions Targeting Diet, Physical Activity, and Weight Gain in Pregnant Women: Systematic Review and Meta-Analysis. JMIR Mhealth Uhealth 2020 Jul 10;8(7):e18255 [FREE Full text] [doi: 10.2196/18255] [Medline: 32673251]
16. Herring SJ, Cruice JF, Bennett GG, Rose MZ, Davey A, Foster GD. Preventing excessive gestational weight gain among African American women: A randomized clinical trial. Obesity (Silver Spring) 2016 Jan;24(1):30-36 [FREE Full text] [doi: 10.1002/oby.21240] [Medline: 26592857]
17. Pollak KI, Alexander SC, Bennett G, Lyna P, Coffman CJ, Bilheimer A, et al. Weight-related SMS texts promoting appropriate pregnancy weight gain: a pilot study. Patient Educ Couns 2014 Nov;97(2):256-260 [FREE Full text] [doi: 10.1016/j.pec.2014.07.030] [Medline: 25153313]
18. Willcox JC, Wilkinson SA, Lappas M, Ball K, Crawford D, McCarthy EA, et al. A mobile health intervention promoting healthy gestational weight gain for women entering pregnancy at a high body mass index: the txt4two pilot randomised controlled trial. BJOG 2017 Oct;124(11):1718-1728. [doi: 10.1111/1471-0528.14552] [Medline: 28220604]
19. Olson CM, Groth SW, Graham ML, Reschke JE, Strawderman MS, Fernandez ID. The effectiveness of an online intervention in preventing excessive gestational weight gain: the e-moms roc randomized controlled trial. BMC Pregnancy Childbirth 2018 May 09;18(1):148 [FREE Full text] [doi: 10.1186/s12884-018-1767-4] [Medline: 29743026]
20. Graham ML, Strawderman MS, Demment M, Olson CM. Does Usage of an eHealth Intervention Reduce the Risk of Excessive Gestational Weight Gain? Secondary Analysis From a Randomized Controlled Trial. J Med Internet Res 2017 Jan 09;19(1):e6 [FREE Full text] [doi: 10.2196/jmir.6644] [Medline: 28069560]
21. Kennelly MA, Ainscough K, Lindsay KL, O Sullivan E, Gibney ER, McCarthy M, et al. Pregnancy Exercise and Nutrition With Smartphone Application Support: A Randomized Controlled Trial. Obstet Gynecol 2018 May;131(5):818-826. [doi: 10.1097/AOG.0000000000002582] [Medline: 29630009]
22. Ainscough KM, O'Brien EC, Lindsay KL, Kennelly MA, O'Sullivan EJ, O'Brien OA, et al. Nutrition, Behavior Change and Physical Activity Outcomes From the PEARS RCT-An mHealth-Supported, Lifestyle Intervention Among Pregnant Women With Overweight and Obesity. Front Endocrinol (Lausanne) 2020 Feb 04;10:938 [FREE Full text] [doi: 10.3389/fendo.2019.00938] [Medline: 32117047]
23. Henriksson P, Sandborg J, Blomberg M, Alexandrou C, Maddison R, Silfvernagel K, et al. A Smartphone App to Promote Healthy Weight Gain, Diet, and Physical Activity During Pregnancy (HealthyMoms): Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2019 Mar 01;8(3):e13011 [FREE Full text] [doi: 10.2196/13011] [Medline: 30821695]
24. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
25. Nordic Council of Ministers. Nordic Nutrition Recommendations 2012. 2012. URL: https://www.livsmedelsverket.se/globalassets/publikationsdatabas/andra-sprak/nordic-nutrition-recommendations-2012.pdf?AspxAutoDetectCookieSupport=1 [accessed 2020-10-27]
26. Josefsson A, Haakstad LA, Bö K. FYSS-kapitel FYSISK AKTIVITET VID GRAVIDITET. Rekommendationer om fysisk aktivitet vid graviditet. 2016. URL: http://www.fyss.se/wp-content/uploads/2017/09/FYSS-kapitel_FA_Graviditet_FINAL_2016-12.pdf [accessed 2020-10-28]
27. Bandura A. Human agency in social cognitive theory. Am Psychol 1989 Sep;44(9):1175-1184. [doi: 10.1037/0003-066x.44.9.1175] [Medline: 2782727]
28. 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]
29. Samdal GB, Eide GE, Barth T, Williams G, Meland E. Effective behaviour change techniques for physical activity and healthy eating in overweight and obese adults; systematic review and meta-regression analyses. Int J Behav Nutr Phys Act 2017 Mar 28;14(1):42 [FREE Full text] [doi: 10.1186/s12966-017-0494-y] [Medline: 28351367]
30. Soltani H, Arden MA, Duxbury AMS, Fair FJ. An Analysis of Behaviour Change Techniques Used in a Sample of Gestational Weight Management Trials. J Pregnancy 2016;2016:1085916 [FREE Full text] [doi: 10.1155/2016/1085916] [Medline: 27034836]
31. Rasmussen KM, Yaktine AL. Weight Gain During Pregnancy: Reexamining the Guidelines. Washington DC: The National Academies Press (US); 2009:71-110.
32. Ruiz JR, Perales M, Pelaez M, Lopez C, Lucia A, Barakat R. Supervised exercise-based intervention to prevent excessive gestational weight gain: a randomized controlled trial. Mayo Clin Proc 2013 Dec;88(12):1388-1397. [doi: 10.1016/j.mayocp.2013.07.020] [Medline: 24290112]
33. Henriksson P, Löf M, Forsum E. Assessment and prediction of thoracic gas volume in pregnant women: an evaluation in relation to body composition assessment using air displacement plethysmography. Br J Nutr 2013 Jan 14;109(1):111-117. [doi: 10.1017/S0007114512000906] [Medline: 22716660]
34. van Raaij JM, Peek ME, Vermaat-Miedema SH, Schonk CM, Hautvast JG. New equations for estimating body fat mass in pregnancy from body density or total body water. Am J Clin Nutr 1988 Jul;48(1):24-29. [doi: 10.1093/ajcn/48.1.24] [Medline: 3389327]

35. Moraeus L, Warensjö Lemming E, Hursti UK, Arnemo M, Sipinen JP, Lindroos A. Riksmaten Adolescents 2016-17: A national dietary survey in Sweden - design, methods, and participation. Food Nutr Res 2018 Aug 28;62 [FREE Full text] [doi: 10.29219/fnr.v62.1381] [Medline: 30186087]
36. Moraeus L, Lindroos AK, Warensjö Lemming E, Mattisson I. Diet diversity score and healthy eating index in relation to diet quality and socio-demographic factors: results from a cross-sectional national dietary survey of Swedish adolescents. Public Health Nutr 2020 Jul;23(10):1754-1765 [FREE Full text] [doi: 10.1017/S1368980019004671] [Medline: 32301415]
37. 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]
38. Migueles JH, Cadenas-Sanchez C, Ekelund U, Delisle Nyström C, Mora-Gonzalez J, Löf M, et al. Accelerometer Data Collection and Processing Criteria to Assess Physical Activity and Other Outcomes: A Systematic Review and Practical Considerations. Sports Med 2017 Sep;47(9):1821-1845. [doi: 10.1007/s40279-017-0716-0] [Medline: 28303543]
39. R [Software]. URL: http://www.R-project.org/ [accessed 2021-02-21]
40. Migueles JH, Rowlands A, Huber F, Sabia S, van Hees V. GGIR: A Research Community–Driven Open Source R Package for Generating Physical Activity and Sleep Outcomes From Multi-Day Raw Accelerometer Data. J Meas Phys Behav 2019;2(3):188-196. [doi: 10.1123/jmpb.2018-0063]
41. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia 1985 Jul;28(7):412-419. [doi: 10.1007/BF00280883] [Medline: 3899825]
42. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med 2011 Feb 20;30(4):377-399. [doi: 10.1002/sim.4067] [Medline: 21225900]
43. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New Jersey: John Wiley and Sons, Inc; Jun 09, 1987.
44. Vickers AJ, Altman DG. Statistics notes: Analysing controlled trials with baseline and follow up measurements. BMJ 2001 Nov 10;323(7321):1123-1124 [FREE Full text] [doi: 10.1136/bmj.323.7321.1123] [Medline: 11701584]
45. Ek A, Alexandrou C, Söderström E, Bergman P, Delisle Nyström C, Direito A, et al. Effectiveness of a 3-Month Mobile Phone-Based Behavior Change Program on Active Transportation and Physical Activity in Adults: Randomized Controlled Trial. JMIR Mhealth Uhealth 2020 Jun 8;8(6). [doi: 10.2196/18531] [Medline: 32510462]
46. Bendtsen M. A Gentle Introduction to the Comparison Between Null Hypothesis Testing and Bayesian Analysis: Reanalysis of Two Randomized Controlled Trials. J Med Internet Res 2018 Oct 24;20(10):e10873 [FREE Full text] [doi: 10.2196/10873] [Medline: 30148453]
47. Bendtsen M. Electronic Screening for Alcohol Use and Brief Intervention by Email for University Students: Reanalysis of Findings From a Randomized Controlled Trial Using a Bayesian Framework. J Med Internet Res 2019 Nov 7;21(11):e14419. [doi: 10.2196/14419] [Medline: 31697242]
48. Bendtsen M. An Electronic Screening and Brief Intervention for Hazardous and Harmful Drinking Among Swedish University Students: Reanalysis of Findings From a Randomized Controlled Trial Using a Bayesian Framework. J Med Internet Res 2019 Dec 17;21(12):e14420. [doi: 10.2196/14420] [Medline: 31845903]
49. Li L, Aris IM, Han WM, Tan KH. A Promising Food-Coaching Intervention Program to Achieve Optimal Gestational Weight Gain in Overweight and Obese Pregnant Women: Pilot Randomized Controlled Trial of a Smartphone App. JMIR Form Res 2019 Oct 24;3(4):e13013 [FREE Full text] [doi: 10.2196/13013] [Medline: 31651407]
50. Hayman M, Reaburn P, Browne M, Vandelanotte C, Alley S, Short CE. Feasibility, acceptability and efficacy of a web-based computer-tailored physical activity intervention for pregnant women - the Fit4Two randomised controlled trial. BMC Pregnancy Childbirth 2017 Mar 23;17(1):96 [FREE Full text] [doi: 10.1186/s12884-017-1277-9] [Medline: 28335767]
51. Choi J, Lee JH, Vittinghoff E, Fukuoka Y. mHealth Physical Activity Intervention: A Randomized Pilot Study in Physically Inactive Pregnant Women. Matern Child Health J 2016 May;20(5):1091-1101. [doi: 10.1007/s10995-015-1895-7] [Medline: 26649879]
52. Walker R, Bennett C, Blumfield M, Gwini S, Ma J, Wang F, et al. Attenuating Pregnancy Weight Gain-What Works and Why: A Systematic Review and Meta-Analysis. Nutrients 2018 Jul 22;10(7):944 [FREE Full text] [doi: 10.3390/nu10070944] [Medline: 30037126]
53. Currie S, Sinclair M, Murphy MH, Madden E, Dunwoody L, Liddle D. Reducing the decline in physical activity during pregnancy: a systematic review of behaviour change interventions. PLoS One 2013 Jun 14;8(6):e66385 [FREE Full text] [doi: 10.1371/journal.pone.0066385] [Medline: 23799096]
54. Statistics Sweden. Population 16-74 years of age by region, highest level of education, age and sex. Year 1985-2019. URL: http://www.statistikdatabasen.scb.se/pxweb/en/ssd/START__UF__UF0506/Utbildning/ [accessed 2020-11-12]
55. Burrows TL, Ho YY, Rollo ME, Collins CE. Validity of Dietary Assessment Methods When Compared to the Method of Doubly Labeled Water: A Systematic Review in Adults. Front Endocrinol (Lausanne) 2019 Dec 17;10:850 [FREE Full text] [doi: 10.3389/fendo.2019.00850] [Medline: 31920966]
56. Bendtsen M, Bendtsen P, Henriksson H, Henriksson P, Müssener U, Thomas K, et al. The Mobile Health Multiple Lifestyle Behavior Interventions Across the Lifespan (MoBILE) Research Program: Protocol for Development, Evaluation, and Implementation. JMIR Res Protoc 2020 Apr 20;9(4):e14894 [FREE Full text] [doi: 10.2196/14894] [Medline: 32310147]

57. Sandborg J, Henriksson P, Larsen E, Lindqvist AK, Rutberg S, Söderström E, et al. Participants' Engagement and Satisfaction With a Smartphone App Intended to Support Healthy Weight Gain, Diet, and Physical Activity During Pregnancy: Qualitative Study Within the HealthyMoms Trial. JMIR mHealth uHealth 2021 Mar 05;9(3):e26159 [FREE Full text] [doi: 10.2196/26159] [Medline: 33666554]

Abbreviations

GWG: gestational weight gain
HOMA-IR: Homeostatic Model Assessment for Insulin Resistance
mHealth: mobile health
MVPA: moderate-to-vigorous physical activity

Edited by G Eysenbach; submitted 27.11.20; peer-reviewed by R Krukowski, AS Aslam, M Hayman, L Redman, R Biviji; comments to author 16.12.20; revised version received 19.01.21; accepted 04.02.21; published 11.03.21.

Please cite as:
Sandborg J, Söderström E, Henriksson P, Bendtsen M, Henström M, Leppänen MH, Maddison R, Migueles JH, Blomberg M, Löf M Effectiveness of a Smartphone App to Promote Healthy Weight Gain, Diet, and Physical Activity During Pregnancy (HealthyMoms): Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(3):e26091
URL: https://mhealth.jmir.org/2021/3/e26091
doi:10.2196/26091
PMID:33704075

Original Paper

A Smartphone App to Restore Optimal Weight (SPAROW) in Women With Recent Gestational Diabetes Mellitus: Randomized Controlled Trial

Karen Lim[1], BSc, MBBS, MRCOG; Shiao-Yng Chan[1], BA (Hons), MBBCHIR, FRCOG, CCT, PhD; Su Lin Lim[2], BSc, PhD; Bee Choo Tai[3], MSc, PhD; Cammy Tsai[4], BSc; Su Ren Wong[4], BSc, PhD; Siew Min Ang[2], MSc; Tong Wei Yew[5,6], MBBS, MRCP; E Shyong Tai[3,5,6], MB ChB, MRCP, FRCP, PhD; Eu Leong Yong[1], MBBS, MRCOG, PhD

[1]Department of Obstetrics and Gynecology, National University Hospital, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[2]Department of Dietetics, National University Hospital, Singapore, Singapore

[3]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore

[4]Department of Rehabilitation, National University Hospital, Singapore, Singapore

[5]Department of Medicine, National University Hospital, Singapore, Singapore

[6]Department of Medicine, Yong Loo Lin School of Medicine, Singapore, Singapore

Corresponding Author:
Eu Leong Yong, MBBS, MRCOG, PhD
Department of Obstetrics and Gynecology, National University Hospital
Yong Loo Lin School of Medicine
National University of Singapore
5 Lower Kent Ridge Rd
Singapore, 119074
Singapore
Phone: 65 81125777
Email: obgyel@nus.edu.sg

Abstract

Background: Women with a history of gestational diabetes mellitus (GDM) are at an increased risk of developing type 2 diabetes mellitus (T2DM). Lifestyle interventions aimed at postpartum weight loss to reduce T2DM risk have been reported, but poor compliance remains a barrier. Smartphone-based interventions may improve compliance, but data on its use in women with recent GDM are limited.

Objective: This trial aimed to investigate the efficacy of a smartphone app in restoring optimal weight following delivery in women with GDM, in the setting of a population with high rates of GDM and type 2 diabetes.

Methods: In this unblinded randomized controlled trial, 200 women with GDM were randomized to receive the intervention or standard care following delivery. The intervention enabled logging of weight, meals, and activity, with web-based interaction with a team comprising dieticians, a physiotherapist, and an occupational therapist. The primary outcome was an achievement of optimal weight (defined as the restoration of first trimester weight if first trimester BMI≤23 kg/m^2 or weight loss of at least 5% from first trimester weight if first trimester BMI>23 kg/m^2) at 4 months post partum. Secondary outcome measures included absolute weight loss, serum metabolic markers, self-reported nutritional intake, health education, and quality of life via questionnaires and user engagement in the intervention group.

Results: In total, 40% (38/96) of women in the intervention group achieved optimal weight at 4 months post delivery compared with 32% (28/93) in the control group (P=.27). Compared with the control group, women in the intervention group reported significantly reduced caloric intake at 4 months after delivery (P<.001) and higher health-directed behavior scores (P=.045). The intervention group also reported increased emotional distress scores (P=.01). At 4 months, participant engagement with the intervention was maintained at 60.8% (SD 33.9%).

Conclusions: Although a statistically significant increase in women achieving healthy weight was not observed, this app remains promising, as women in the intervention group reported improved health behaviors and lower caloric intake. Importantly, the

high retention rates suggest that a larger study with a longer follow-up period might confirm the effectiveness of this app for weight management.

Trial Registration: ClinicalTrials.gov NCT03324737; https://clinicaltrials.gov/ct2/show/NCT03324737

International Registered Report Identifier (IRRID): RR2-10.1186/s12889-019-7691-3

(JMIR Mhealth Uhealth 2021;9(3):e22147) doi:10.2196/22147

KEYWORDS

randomized controlled trial; gestational diabetes mellitus; prevention; weight loss; mobile phone

Introduction

Background

Type 2 diabetes mellitus (T2DM) is an important cause of morbidity and mortality, estimated to have affected 415 million people worldwide in 2015 [1]. Women with gestational diabetes mellitus (GDM) in their pregnancies have up to seven-fold increased risk of developing subsequent T2DM [2]. The postpartum period following GDM represents a unique opportunity for early intervention to lower the risk of subsequent T2DM, possibly delaying the point at which prediabetes or T2DM is detected through health screening or other clinical encounters at later stages in their lives.

Weight gain following pregnancy has been shown to increase T2DM risk in women with a history of GDM [3]. A post hoc analysis of women with a history of GDM recruited in the diabetes prevention program showed that weight loss through intensive lifestyle modification reduced T2DM risk by 53% [4]. Although there is often ready acceptance of lifestyle changes among women during the antenatal period after the diagnosis of GDM [5], adherence to these lifestyle changes after delivery remains to be a challenge. This is especially so in the early postnatal period, where the challenges of recovering from the delivery process, breastfeeding, and caring for a newborn all lead to a chronic lack of sleep and fatigue [6,7].

Telephone-based [8] and face-to-face [9,10] interventions have been shown to be moderately effective in reducing postpartum weight retention in women with GDM, but poor compliance due to time constraints is a constantly cited barrier to successful postpartum weight loss [9,11]. In addition, these methods are labor-intensive and may depend on the busy mother being available at definite time slots. Questionnaires indicate that women with GDM in Singapore prefer web-based health education or apps for health education [12], which is consistent with findings from a trial that showed improved postpartum weight loss in women with GDM using a web-based intervention [13]. App-based interventions appear to be effective in improving nutritional behaviors and weight loss in the general population [14], but the literature on app-based interventions for weight loss during or after pregnancy is limited [15-20]. Reduced gestational weight gain has been reported with the use of an app in the antenatal period [17,19], but a positive effect on postpartum weight retention has yet to be demonstrated [18]. To date, there have been no studies on the use of apps for postpartum weight loss in women with a history of GDM.

Singapore has one of the highest rates of GDM globally, reaching up to 25% of pregnancies [21], as well as one of the highest smartphone penetration rates in Asia (approximately 90%) [22], both of which contribute to the clinical relevance and feasibility of this trial in this population.

Objectives

The Smartphone App to Restore Optimum Weight (SPAROW) trial was a randomized controlled trial (RCT) undertaken to examine the effectiveness of an interactive app to restore optimal weight in postnatal women with a recent history of GDM. We measured the ability of the app to achieve first trimester weight in subjects at 4 months post partum if their first trimester BMI was within the normal range or weight loss of at least 5% of first trimester pregnancy weight if they were overweight. Other outcomes examined were the effects of the app on dietary and health education parameters and on a series of cardiometabolic markers. Our hypothesis was that a smartphone app would be effective in achieving optimal weight in women with GDM.

Methods

Study Design

The detailed design and methodology of the SPAROW trial are described elsewhere [12]. In brief, participants were recruited from women who had recently delivered at the National University Hospital (NUH), Singapore. The trial obtained ethics approval from the Domain Specific Review Board and was registered at ClinicalTrials.gov on October 30, 2017 (identifier: NCT03324737).

Recruitment

The electronic medical records of all women in the postnatal ward were screened daily by the research members to determine eligibility. If eligibility criteria were met, women were approached by a member of the study team in the ward before discharge, and written informed consent was obtained.

Participants

Eligible women included postpartum women aged ≥21 years diagnosed with GDM antenatally (between 24-34 weeks of gestation) using a 75-g 3 time point oral glucose tolerance test (OGTT) according to the 2013 World Health Organization criteria (fasting plasma glucose ≥5.1 mmol/L, 1-hour plasma glucose ≥10.0 mmol/L, and/or 2-hour plasma glucose ≥8.5 mmol/L) [23]. Women were required to own a smartphone and be able to use an app independently**.** In addition, the first trimester weight must have been documented at or before 13 weeks. Women with pre-existing type 1 diabetes mellitus or T2DM and women who delivered before 36 weeks were excluded.

Randomization and Enrollment

Participants were randomized at the recruitment visit to the intervention or control arm using a random permuted block design with a block size of 4. An independent researcher generated the set of sequences and assigned participants to the intervention or control groups using sequentially numbered sealed opaque envelopes to ensure allocation concealment until the interventions were assigned. Owing to the nature of the intervention, blinding of participants and assessors was not possible**.**

Standard Care Group (Control)

Women in the standard care arm were given a follow-up appointment at 6 weeks post partum for review by a clinician with a routine postnatal check where dietary advice was provided and a repeat OGTT was performed. Women found to have impaired fasting glucose or impaired glucose tolerance were issued a letter reinforcing healthy lifestyle changes and encouraged to consult a family physician. Participants with T2DM were referred to their family physician or internist of their choice. Women with a normal OGTT were informed of their normal results.

App Group (Intervention)

Women allocated to the intervention arm were instructed to download an app called *Nutritionist Buddy (nBuddy)* immediately post partum before discharge and were briefed on its use by a member of the study team. During this briefing, participants were taught how to use the different functions of the app, and the live chat function was demonstrated. A colored information booklet was also provided to the participant on how to use the app. The nBuddy app is locally developed and was based on the Obesity-Related Behavioral Intervention Trials model for developing behavioral treatments for chronic disease [24]. This has been described in further detail in our protocol [12].

Through the app, participants were encouraged to work toward personalized weight targets. Calorie and activity level goals set to attain the target weight were individualized according to participants' profiles and were automatically generated by the app. Breastfeeding status was factored into the eventual recommended caloric goal and adjusted accordingly if women stopped breastfeeding. If participants had difficulty achieving goals that were initially set, the targets were readjusted by members of the study team in agreement with the participants. Once the initial target weight was achieved, new goals were set to attain the final target weight.

Participants were advised to log their food intake daily by selecting a pre-entered meal from the app's nutritional database of more than 11,000 locally available foods. Reminder notifications were sent 2 hours after the standard meal times if participants failed to do so. Prompts recommending healthier food alternatives catered to ethnicity were activated automatically if foods that were unhealthy or high in calories were selected or if participants had exceeded their calorie limit for the day. Optional prompts for meal planning could be selected by participants, which would suggest healthy food options 1 hour before meal time.

The step-counting function allowed users to track their physical activities. Participants were also able to manually input a range of physical activities, which were then automatically translated into step counts.

A total of 16 video clips covering themes such as diet, exercise, and emotional health for new mothers, as well as the benefits of breastfeeding and weaning diet for babies, were specially produced for the study. These clips were short (3 minute) to cater to time-challenged mothers and were made available for viewing at appropriate times after delivery [12].

A unique feature of the intervention was that it allowed live interaction between the participants and the study team consisting of dietitians, physiotherapists, and occupational therapists. All members of the research team attended a 1-hour training session conducted by the chief dietician on how to support the participants through the app. After the app was downloaded by the participant, the study team member who recruited the participant in the ward initiated the first live chat on the same day. Following this, team members actively engaged participants on the web for 5 minutes to 10 minutes daily during weekdays for the first week, followed by 2 to 3 times a week in subsequent weeks. Questions raised by participants throughout the week were responded to within a working day. Issues faced by the participants were discussed between team members to ensure the provision of appropriate and relevant advice. In addition, adjustment strategies and self-care behaviors were encouraged to aid participants in adapting to motherhood and healthy habit formation.

Data Collection

The first trimester weight was defined as the weight taken before 13 weeks of gestation. If multiple weights were available in the first trimester, the mean of the weights was taken. Participants in the control and intervention arms were scheduled for follow-up visits at 6 weeks and 4 months postnatally, during which a series of investigations were performed [12]. A calibrated digital weighing machine (Seca 799) was used for all clinic-recorded weights, with an accuracy of 50 g for <50 kg and 100 g for 50-150 kg.

Data on app usage by participants were manually retrieved via the app's backend dashboard in combination with usage statistics obtained from the app developer. We tracked engagement data in the following categories: (1) meal logging, (2) step logging, (3) weight logging, and (4) interactivity with health coaches. Owing to the lack of available app engagement data in the current literature, we defined adequate engagement as logging of at least one meal per day for more than 50% of the month, logging or syncing of steps at least 8 times per month, logging of weight at least 4 times per month, and communicating with coaches on the app at least 8 times per month. The overall utilization rate represented the percentage of days in a month in which an app component was engaged. To standardize calculations, it was assumed that each month had a total of 30 days.

Data were also extracted from the step count function of the app for the intervention group at baseline, 6 weeks, and 4 months.

Three-day food diaries were mailed to the participants 2 weeks before their follow-up visits. They were advised to input meals consumed on 2 weekdays and 1 weekend day and submit them at the follow-up visit.

The Health Education Impact Questionnaire (heiQ), RAND-12, and self-efficacy questionnaires were self-administered at the follow-up visits. The heiQ questionnaire is a well-validated tool used to assess self-management support and patient education programs [25] and for evaluating interventions for chronic conditions such as T2DM [26-28]. It is a self-reported questionnaire with 42 items divided into 8 subscales. Higher scores on each subscale reflect higher levels of skill, knowledge, or understanding, with the exception of the emotional distress subscale, which is reversed, and higher scores indicate higher levels of distress. The RAND-12 is a well-validated subjective measure of quality of life, developed from the Short-Form Health Survey (12 items) [29]. It is a self-administered global questionnaire with 12 questions that are grouped into 2 domains: physical and mental health component scores. The Self-Efficacy to Regulate Exercise questionnaire assesses participants' confidence in their ability to exercise regularly in specific circumstances [30].

Study Outcomes

The primary outcome measure of this study was the percentage of women who were able to achieve their first trimester weight at 4 months post partum if their first trimester BMI was ≤23 kg/m^2 or weight loss of at least 5% with respect to first trimester weight if their first trimester BMI was >23 kg/m^2.

Prespecified secondary outcomes were as follows:

1. A 75 g 2-hour OGTT, glycated hemoglobin (HbA_{1c}), C-peptide, homeostatis model assessment of insulin resistance, lipid profiles, liver function, high-sensitivity C-reactive protein, and interleukin-6.
2. Mean absolute weight loss.
3. Breastfeeding status.
4. Blood pressure.
5. Right hand grip strength and waist circumference.
6. heiQ, self-efficacy, and RAND-12 questionnaire.
7. Caloric and macronutrient intake assessed by a 3-day food diary.

Statistical Analysis and Sample Size

The sample size was estimated based on the primary outcome of attaining optimal weight at 4 months post delivery. This was defined as their first trimester weight if their first trimester BMI was ≤23 kg/m^2 or weight loss of at least 5% if their first trimester BMI was >23 kg/m^2. We estimated that 20% of women in the control group would attain their target weight based on similar studies [8]. An additional 20% of women attaining their target weight in the intervention group was thought to be clinically significant, corresponding to 40% of women receiving the intervention. We calculated that 75 individuals in each group would be required to achieve a power of at least 80%. Accounting for an approximate attrition rate of 15%, we aimed to recruit 100 individuals in each group.

For the analysis of the outcomes based on weight restoration and breastfeeding status (exclusive, partial, or none), the Pearson χ^2 test was used to compare the respective proportions between groups at 4 months post partum. The effect estimate was expressed as an odds ratio (OR) with a 95% CI. Subsequent analyses via the mixed effect logistic regression model were used to adjust for ethnicity, parity, and the effect of time, taking into account repeated measures at 6 weeks and 4 months.

For secondary outcomes involving absolute weight change, anthropometric measurements, diet and nutrition markers as well as RAND-12, self-efficacy, and heiQ domains at 4 months post partum, the mean difference between the intervention and control groups was first compared based on a two tailed *t* test. Furthermore, adjusted analyses were performed using the linear mixed model to adjust for the respective baseline covariate (where available) and the effect of time, taking into account possible intrasubject correlation between the repeated outcomes at 6 weeks and 4 months.

All analyses were performed on an intention-to-treat (ITT) basis, assuming a two-sided test at the 5% level of significance using StataCorp 2019 (StataCorp LLC).

Results

Study Population

Women with GDM (n=773) who delivered at NUH, Singapore, between November 2017 and February 2019 were screened for eligibility (Figure 1). A total of 346 women did not meet the eligibility criteria, and 227 declined participation. In total, 200 women consented to participate in the trial and were randomized either to the intervention arm (n=101) or the control arm (n=99). At the end of the study, 95% of those recruited (96 in intervention and 93 in control) provided information on the outcome analyses for the duration that they were observed. In total, 11 patients (5 intervention and 6 control) were lost to follow-up at week 6 and a further 7 (1 intervention and 6 control) were lost to follow-up at month 4.

The characteristics of the participants who contributed to the outcome analyses are shown in Table 1. The baseline characteristics were largely comparable between the control and intervention groups. The mean gestational weight gain did not differ significantly between the groups.

Figure 1. Trial flow. GDM: gestational diabetes mellitus; NUH: National University Hospital; OGTT: oral glucose tolerance test; WHO: World Health Organization.

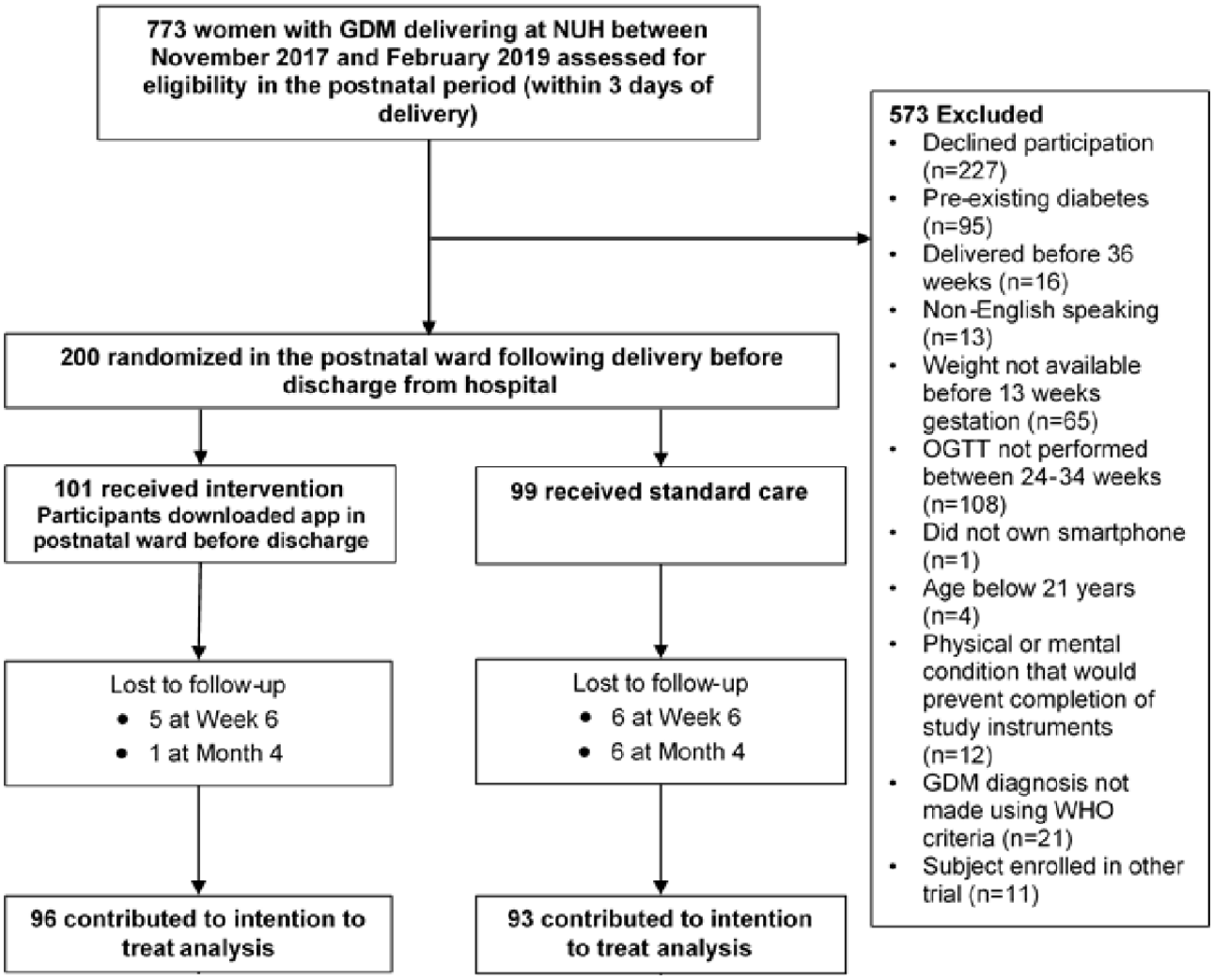


Table 1. Demographics and clinical characteristics of trial participants.

Characteristics	Intervention (n=96)	Control (n=93)	Total (N=189)
Age (years), mean (SD)	32.6 (4.5)	32.4 (4.2)	32.5 (4.3)
Ethnicity, n (%)			
Chinese	46 (47.9)	54 (58.1)	100 (52.9)
Malay	35 (36.5)	22 (23.7)	57 (30.2)
Indian	11 (11.5)	13 (14.0)	24 (12.7)
Others	4 (4.2)	4 (4.3)	8 (4.2)
Height at delivery (m), mean (SD)	1.58 (0.06)	1.60 (0.05)	1.59 (0.06)
First trimester weight (kg), mean (SD)	63.1 (16.0)	61.9 (11.3)	62.5 (13.9)
First trimester BMI (kg/m^2), median (IQR)	24.6 (20.7-27.8)	23.4 (21.0-27.0)	23.7 (20.7-27.6)
BMI at delivery (kg/m^2), median (IQR)	26.9 (23.0-29.5)	25.8 (22.8-28.6)	26.2 (22.9-28.9)
Blood pressure at recruitment (mm Hg), mean (SD)			
Systolic	107.1 (11.6)	106.8 (10.7)	106.9 (11.2)
Diastolic	63.7 (9.3)	63.3 (9.9)	63.5 (9.6)
Parity, n (%)			
1	39 (40.6)	48 (51.6)	87 (46.0)
2	38 (39.6)	33 (35.5)	71 (37.6)
3+	19 (19.8)	12 (12.9)	31 (16.4)
Level of education, n (%)			
Secondary	11 (11.5)	6 (6.5)	17 (9.0)
GCE[a] A level or IB[b] or ITE[c] or NTC[d]	5 (5.2)	6 (6.5)	11 (5.8)
Diploma or advanced diploma or polytechnic	21 (21.9)	19 (20.4)	40 (21.2)
Bachelor's degree	49 (51.0)	40 (43.0)	89 (47.1)
Postgraduate	10 (10.4)	22 (23.7)	32 (16.9)
Employed, n (%)	74 (77.1)	80 (86.0)	154 (81.5)
History of previous cesarean section, n (%)	22 (23.7)	20 (22.5)	42 (23.1)
Pre-existing hypertension, n (%)	1 (1.0)	1 (1.1)	2 (1.1)
History of gestational diabetes in previous pregnancies, n (%)	13 (13.5)	13 (14.0)	26 (13.8)
History of impaired fasting glucose or glucose tolerance, n (%)	0 (0.0)	1 (1.1)	1 (0.5)
Oral glucose tolerance test at diagnosis (mmol/L), mean (SD)			
Fasting	4.6 (0.4)	4.7 (0.5)	4.6 (0.5)
1 hour	9.8 (1.4)	10.2 (1.4)	10.0 (1.4)
2 hours	8.4 (1.4)	8.7 (1.2)	8.5 (1.3)
Gestational age at gestational diabetes diagnosis (weeks), mean (SD)	27.1 (2.0)	27.0 (1.7)	27.1 (1.8)
HbA_{1c}[e] at diagnosis (%), mean (SD)[f]	5.4 (0.3)	5.2 (0.4)	5.3 (0.4)
Gestational weight gain (kg), mean (SD)	8.84 (4.2)	9.19 (4.7)	9.0 (4.5)

[a]GCE: General Certificate of Education.

[b]IB: International Baccalaureate.

[c]ITE: Institute for Technical Education.

[d]NTC: National Technical Certificate.

[e]HbA_{1c}: glycated hemoglobin.

[f]12 observations with missing HbA_{1c} measurements at diagnosis.

Primary Outcome

In the ITT analysis, 40% (38/96) of women in the intervention arm achieved optimal weight at 4 months compared with 30% (28/93) in the control arm (OR 1.40, 95% CI 0.76 to 2.58). Adjustment for ethnicity, parity, and the effect of time via the linear mixed model did not materially alter the results (OR 1.55, 95% CI 0.53 to 4.54).

Secondary Outcomes

Although women in the intervention arm achieved a mean difference in weight reduction of 1.05 kg compared with the control arm (95% CI 0.14 to 2.24; *P*=.08), this difference was not statistically significant. There were no significant differences in anthropometric measurements, breastfeeding status, and OGTT results at 4 months (Table 2). Metabolic serum markers did not differ between the groups at 4 months, except for lower HbA_{1c} levels in the control group at 4 months (Multimedia Appendix 1). At 4 months postnatally, 3% (3/87) and 0% (0/95) had impaired fasting glucose and 14% (12/87) and 19% (18/95) had impaired glucose tolerance in the intervention and control groups, respectively. None of the participants had developed T2DM at 6 weeks or 4 months.

Table 2. Mean anthropometric measurements in the intervention and control group and their mean differences at month 4[a].

Outcome	Intervention (n=96)	Control (n=93)	Mean difference (95% CI)	*P* value
Absolute difference in weight (kg) from first trimester				
Unadjusted	−1.39	−0.34	−1.05 (−2.24 to 0.14)	.08
Adjusted	−0.63	0.08	−0.71 (−1.76 to 0.33)	.18
Waist circumference (cm)				
Unadjusted	83.7	85.0	−1.26 (−4.41 to 1.89)	.43
Adjusted	84.1	85.6	−1.46 (−4.42 to 1.52)	.34
Systolic blood pressure (mm Hg)				
Unadjusted	106.8	106.7	0.06 (−3.09 to 3.22)	.97
Adjusted	107.0	106.7	0.23 (−2.41 to 2.88)	.86
Right hand grip strength (kg)				
Unadjusted	22.46	23.36	−0.90 (−2.30 to 0.50)	.21
Adjusted	22.40	22.84	−0.44 (−1.77 to 0.89)	.52
Exclusive breastfeeding (days)				
Unadjusted	51.6	52.9	0.95 (0.53 to 1.70)	.86
Adjusted	50.5	48.3	1.26 (0.41 to 3.86)	.68
Fasting OGTT[b] test (mmol/L)				
Unadjusted	4.77	4.75	0.02 (−0.12 to 0.16)	.76
Adjusted	4.73	4.74	−0.02 (−0.12 to 0.08)	.73
2-hour OGTT (mmol/L)				
Unadjusted	6.59	6.65	−0.06 (−0.56 to 0.43)	.80
Adjusted	6.63	6.58	0.05 (−0.39 to 0.48)	.83

[a]Adjusted analysis is based on a linear mixed effect model adjusting for the respective baseline covariate (in addition, ethnicity and parity were also included for exclusive breastfeeding) and the effect of time, taking into account possible intrasubject correlation between the repeated outcomes at week 6 and month 4.

[b]OGTT: oral glucose tolerance test.

Diet and Nutrition

At 6 weeks, women in the intervention group reported reduced total caloric intake (−591.3 kcal, 95% CI −717.1 to −465.4), total fat (−25.3 g, 95% CI −31.7 to −19.0), protein (−24.1 g, 95% CI −31.1 to −17.0), carbohydrate (−67.4 g, 95% CI −84.6 to −50.2), and sugar (−22.9 g, 95% CI −30.0 to −15.7) compared with controls. Similarly, at 4 months, the intervention group reported reduced total caloric intake (−614.2 kcal, 95% CI −751.5 to −476.9), total fat (−20.7 g, 95% CI −27.2 to −14.2), protein (−25.7 g, 95% CI −33.1 to −18.3), carbohydrate (−71.2 g, 95% CI −90.3 to −52.1), and sugar (−27.9 g, 95% CI −35.7 to −20.1; Table 3). The results remained similar even after adjusting for baseline covariates and the effect of time. The relative contribution of sugar to overall caloric intake was significantly lower in the intervention arm, with an adjusted mean difference of −0.019 (95% CI −0.028 to −0.010; *P*<.001; Table S2 in Multimedia Appendix 1). The relative contributions of fat, protein, and carbohydrate did not differ significantly between the groups, suggesting that a reduction in sugar intake was the likely driver for reduced overall caloric intake in the intervention arm.

Table 3. Mean nutritional intake in the intervention and control and their mean differences at month 4[a].

Diet and nutrition markers	Intervention (n=95)[b]	Control (n=87)[b]	Mean difference (95% CI)	*P* value
Calories (kcal)				
Unadjusted	1441.3	2055.5	−614.2 (−753.3 to −475.2)	<.001
Adjusted	1463.2	2065.9	−602.7 (−717.6 to −487.9)	<.001
Total fat (g)				
Unadjusted	70.7	91.4	−20.7 (−27.4 to −14.2)	<.001
Adjusted	64.0	87.0	−23.0 (−28.3 to −17.8)	<.001
Protein (g)				
Unadjusted	56.1	81.8	−25.7 (−33.1 to −18.3)	<.001
Adjusted	64.4	89.3	−24.9 (−30.8 to −19.0)	<.001
Carbohydrate (g)				
Unadjusted	168.5	239.7	−71.2 (−90.3 to −52.1)	<.001
Adjusted	169.2	238.5	−69.3 (−84.9 to −53.7)	<.001
Sugar (g)				
Unadjusted	37.8	65.7	−27.9 (−35.7 to −20.1)	<.001
Adjusted	38.6	64.0	−25.4 (−31.3 to −19.5)	<.001

[a]Adjusted analysis based on a linear mixed effect model adjusting for the respective baseline covariate and the effect of time, taking into account possible intrasubject correlation between repeated outcomes at week 6 and month 4.

[b]In addition to the 11 women who were lost to follow-up at week 6, diet and nutrition markers were not available at week 6 for 7 women (1 intervention and 6 control).

Physical Activity

Participants in the intervention group logged a mean step count of 4065 (95% CI 3483.6 to 4645.7) at 6 weeks. The mean step count increased to 4880 (95% CI 4195.4 to 5565.5) at 4 months (*P*=.04).

Health Behaviors

Women in the intervention group reported higher health-directed behavior scores on the heiQ questionnaire (mean difference 0.16, 95% CI 0.004 to 0.32) than controls at 4 months (Table S3 in Multimedia Appendix 1). The health-directed behavior component of the heiQ questionnaire assesses changes in lifestyle through healthy behaviors such as diet, exercise, and relaxation routines aimed at either disease prevention and/or health promotion. High scores indicated high levels of healthy behaviors [25]. Specific questions that comprise the health-directed behavior component of the heiQ are presented in Table 4.

Table 4. Percentage of "agree" and "strongly agree" responses to questions in the health-directed behavior component of the Health Education Impact Questionnaire in the intervention group.

Question	Week 6, n (%)	Month 4, n (%)
On most days of the week, I do at least one activity to improve my health (eg, walking, relaxation, exercise)	85 (84.2)	95 (90.5)
I do at least one type of physical activity every day for at least 30 min (eg, walking, gardening, housework, golf, bowls, dancing, Tai Chi, swimming)	75 (74.3)	95 (80.0)
On most days of the week, I set aside time for healthy activities (eg, walking, relaxation, exercise)	67 (67.7)	65 (68.4)
I walk for exercise, for at least 15 min per day, most days of the week	83 (82.2)	82 (86.3)

However, increased emotional distress scores were observed in the intervention group (0.21, 95% CI 0.05 to 0.38). Women in the intervention group also reported lower scores on the physical component summary of the RAND-12 questionnaire (*P*=.04), but this difference was no longer statistically significant after adjustment (Table S3 in Multimedia Appendix 1).

At 4 months, women in the intervention group reported higher scores in question 6 (*P*=.02, unadjusted value) and 11 (*P*=.04, unadjusted value) of self-efficacy to regulate the exercise questionnaire. This addressed how confident participants felt in being able to perform their exercise routine regularly (3 or more times a week) despite physical discomfort experienced during exercise (question 6) and when they had other time commitments (question 11; Table S4 in Multimedia Appendix 1).

User Engagement

The participants' engagement with the app is presented in Table 5. On average, the usage of at least one component of the app by participants was 70.8% in month 1 and 60.8% in month 4.

In the first month, 62.4% (63/101), 83.2% (84/101), and 78.2% (79/101) of users logged their meals, step counts, and weight (as defined earlier), which dropped to 41.6%, 76%, and 61.4% at month 4, respectively.

Table 5. Participant engagement with the app in the intervention group.

App usage by participants (n=101)	Month 1	Month 2	Month 3	Month 4	4-month average
Overall utilization rate (Percentage of days an app component was used per user), mean (SD)	70.8 (30.6)	69.3 (30.9)	61.3 (34.0)	60.8 (33.9)	65.5 (29.0)
Percentage of users who logged meals (≥15 times a month), n (%)	63 (62.4)	63 (62.4)	50 (49.)5	42 (41.6)	54 (54.0)
Percentage of users with step count log (≥8 times a month), n (%)	84 (83.2)	85 (84.2)	78 (77.2)	77 (76.2)	81 (80.2)
Percentage of users with weight log (≥4 times a month), n (%)	79 (78.2)	71 (70.3)	60 (60.4)	62 (61.4)	68 (67.6)
Percentage of users with interactive exchange (≥8 times a month), n (%)	33 (32.7)	31 (30.7)	21 (20.8)	20 (19.8)	26 (25.7)

Discussion

Overview of the Findings

In this RCT involving an app-based lifestyle intervention among postpartum women with recent GDM, the intervention did not achieve a statistically significant difference in the primary outcome. However, the intervention was effective in promoting an overall healthier lifestyle with reduced self-reported caloric intake and improved health-directed behaviors in women with a history of GDM. Utilization of the app remained to be constantly high throughout the study period, with 60% usage of the intervention at 4 months.

The average utilization rate in postnatal women appears to be comparable with populations utilizing apps for chronic disease management, which reported nonengagement rates as high as 40% in a recent review [31]. Across postnatal lifestyle interventions, the engagement level in this app-based study was significantly higher than in studies relying on telephone-based and face-to-face interventions, which reported full attendance rates of only 10% to 15% [8,9]. Convenience and portability may be key factors associated with greater app engagement in time-pressed postnatal women, which in turn may lead to better recommendation adherence and outcomes [32]. A pilot study investigating the use of a smartphone intervention for postpartum weight loss in 40 women who were mostly overweight or obese showed no significant improvement in weight loss at 16 weeks. However, post hoc analysis showed a significant difference in women who were compliant (>70% adherence) to the intervention, suggesting a possible benefit with increased compliance [18]. Our study provides support for this easily scalable app intervention to be applied to larger numbers of subjects to assess its ability to achieve significant weight loss.

A possible reason for our trial being unable to detect a clear improvement in achievement of optimal weight, the primary outcome, was likely due to a higher proportion of women achieving optimal weight in the control arm of 32.2% than the expected 20% based on other studies [8]—the postulated proportion that was used to estimate the sample size for this trial. The continued application of dietary advice obtained from regular dietitian reviews in the antenatal period could have contributed to the higher than expected rates of weight loss for women in the control arm. Another possible explanation for this difference is the Hawthorne effect, wherein women in the control arm might be more likely to modify their health behaviors in response to their awareness of being observed, resulting in social desirability bias. In addition, women who agreed to participate in the study were also likely to represent a highly motivated group of individuals compared with the general population, which could explain the higher rate of optimal weight achievement in the control arm.

The ethnic makeup in the 2 arms differed slightly, with fewer Chinese and more Malays in the intervention arm. Although Malay and Indian people are known to have greater metabolic risk and adiposity [33], adjustment for ethnicity did not alter the results substantially; therefore, ethnicity is unlikely to be a major contributor to our results.

Women who received the intervention reported significantly reduced caloric intake by over 600 calories, which was mainly due to a reduction in sugar intake in the intervention group. Although this finding needs to be interpreted with caution because of inherent limitations of self-reported food diary studies, our findings are consistent with a meta-analysis of data from 41 studies by Villinger et al [14] showing that app-based mobile interventions are effective for changing nutrition behaviors in a wide range of study settings.

Women who received the intervention reported improved health-related behaviors, as evidenced by the statistically significant increase in scores in the health-related behavior domain of the heiQ questionnaire. Similarly, a study investigating the use of a mobile health intervention among people with T2DM showed higher scores in the skill and technique acquisition domain of the heiQ questionnaire after 1 year [34]. Women in the intervention also reported significantly higher scores in 2 questions of the self-efficacy to exercise questionnaire, which assessed their ability to persist with regular exercise despite physical discomfort and time constraints. This suggests that the intervention may improve health behaviors, which could possibly translate into significant weight loss over a longer period of time.

Increased step counts between week 6 and month 4 were noted in the intervention group. The overall step counts, although increased, were modest compared with those in other studies, which achieved up to 10,000 steps per day [35]. However, step counts were obtained from the step count function of the

smartphone rather than a pedometer, and hence may be an underrepresentation of the actual number of steps taken.

The intervention was unique in that it had a large databank of popular Southeast Asian street foods and had built-in prompts to suggest healthier alternatives that were culturally appropriate, an important feature in Singapore's multiethnic society. Its interactive features enabled real-time inputs to inculcate adjustment strategies and self-care behaviors, as well as specially commissioned videos relevant to the recently parturient mother in Asian societies.

Women receiving the intervention reported higher emotional distress scores than those receiving the intervention. Lower scores in the physical component summary but not the mental component summary of the RAND-12 questionnaire in the intervention group may suggest that the source of distress was related to physical fitness rather than emotional problems. These findings may be explained by an increased awareness of potentially developing T2DM and the demands of adhering to lifestyle and dietary modifications to reduce this risk—all of which may increase emotional distress in women receiving the intervention in the short term. These findings differ from trials on the use of an app in the antenatal period to reduce gestational weight gain, which showed no change in mood and quality of life with the intervention [36]. However, the study also showed that women reported higher depression and lower physical quality of life scores during pregnancy than during the postpartum period. This may reflect the different sources of background stressors during the different stages of the pregnancy journey, which may moderate the interaction between the intervention and quality of life.

Although there are several trials investigating the use of lifestyle interventions to improve postpartum weight loss in women with and without GDM [8,13,37,38], this is the first RCT investigating the use of an app for postpartum weight loss in women with recent GDM. Our trial demonstrated that an app supported by interactions with a group of health care professionals may improve health behaviors in postpartum women and sustain high user engagement. This resulted in reduced caloric intake and improved health behaviors when assessed by self-report. As such apps for diet and nutritional intake tracking are currently available globally, implementation of similar studies in other centers would be feasible. In addition, our study population likely represents the ethnic groups of many Southeast Asian countries in which women are known to have a high prevalence of GDM and T2DM.

Limitations

The main limitation of our study was that it was underpowered to detect a statistically significant difference in the primary outcome, likely due to the reasons mentioned earlier. Another limitation is that the outcomes in which significant improvements were detected (caloric-intake and health-directed behaviors) were self-reported rather than objectively measured outcomes. Studies have shown that self-reported physical activity is often an overestimation of actual physical activity [39], which could also explain the lack of significant improvement in weight loss in the intervention group. This emphasizes the need for objectively measured physical activity in future studies, possibly in the form of pedometers provided to all participants in the study. We also acknowledge the limitations of interpreting decreased caloric intake without concomitant data on energy expenditure, as reduced caloric intake may be related to reduced physical activity during the postnatal period. However, as changes in body weight and breastfeeding may be on the causal pathway between the intervention and caloric intake, it is not appropriate to adjust for these factors in our regression model. Similarly, controlling for overall caloric intake when comparing differences in specific macronutrients may not be appropriate, as overall caloric intake lies in the causal pathway between intervention and macronutrient intake. A separate modeling study to investigate the causal pathways involved will be conducted in the future.

The challenge with any intervention remains its ability to sustain long-term maintenance of optimal weight, as positive effects on weight loss have been shown to be negated following the cessation of the intervention [8]. However, there is some evidence to suggest that long-term follow-up following an intervention may lead to sustained weight loss up to 13 years postintervention [40]. This implies that if women continue with the intervention for long enough, healthier food and lifestyle choices may become routine, which may be key in sustaining optimal weight and reducing T2DM risk in these women. Given the apparent acceptability of this app intervention with relatively high and sustained engagement rates, it is possible that long-term follow-up of these women would demonstrate even greater weight loss or weight maintenance, which is a crucial step in the battle against the growing T2DM epidemic. In addition, app-based interventions would be particularly valuable in the delivery of care during the current COVID-19 pandemic and its aftermath, which will rely heavily on telemedicine to minimize direct patient contact.

Conclusions

This RCT suggests that an app may improve health behaviors and can maintain high user engagement. The results of this study support the need for a larger RCT with greater statistical power and longer-term follow-up to track long-term weight loss and cardiometabolic outcomes.

Acknowledgments

The authors would like to thank Ms Jean Ho for help with editing the manuscript. Funding for this study was obtained from a Health Services Research Grant from the National Medical Research Council, Ministry of Health, Singapore, granted on March 16, 2017. This funding source had no role in the design of this study and did not have any role during its execution, analyses, interpretation of the data, or decision to submit results.

Authors' Contributions

EL conceived the study. EL, CC, SY, SL, TW, ES, and KL initiated the study design, and KL, SL, SM, CT, and SR helped with implementation. EL is a grant holder. BC assisted in the analysis and interpretation of data. All authors contributed to the refinement of the study protocol and have approved the final manuscript.

Conflicts of Interest

SY is part of an academic consortium that has received grants from Nestlé S.A. outside the submitted work.

Multimedia Appendix 1
Supplementary Tables.
[DOCX File , 22 KB - mhealth_v9i3e22147_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V1.6.1).
[PDF File (Adobe PDF File), 1217 KB - mhealth_v9i3e22147_app2.pdf]

References

1. Ogurtsova K, da Rocha Fernandes J, Huang Y, Linnenkamp U, Guariguata L, Cho N, et al. IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040. Diabetes Research and Clinical Practice 2017 Jun;128:40-50. [doi: 10.1016/j.diabres.2017.03.024]
2. Bellamy L, Casas J, Hingorani AD, Williams D. Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis. The Lancet 2009 May;373(9677):1773-1779. [doi: 10.1016/s0140-6736(09)60731-5]
3. Bao W, Yeung E, Tobias DK, Hu FB, Vaag AA, Chavarro JE, et al. Long-term risk of type 2 diabetes mellitus in relation to BMI and weight change among women with a history of gestational diabetes mellitus: a prospective cohort study. Diabetologia 2015 Mar 22;58(6):1212-1219. [doi: 10.1007/s00125-015-3537-4]
4. Aroda VR, Christophi CA, Edelstein SL, Zhang P, Herman WH, Barrett-Connor E, et al. The Effect of Lifestyle Intervention and Metformin on Preventing or Delaying Diabetes Among Women With and Without Gestational Diabetes: The Diabetes Prevention Program Outcomes Study 10-Year Follow-Up. The Journal of Clinical Endocrinology & Metabolism 2015 Apr;100(4):1646-1653. [doi: 10.1210/jc.2014-3761]
5. Stage E, Ronneby H, Damm P. Lifestyle change after gestational diabetes. Diabetes Research and Clinical Practice 2004 Jan;63(1):67-72. [doi: 10.1016/j.diabres.2003.08.009]
6. Lie MLS, Hayes L, Lewis-Barned NJ, May C, White M, Bell R. Preventing Type 2 diabetes after gestational diabetes: women's experiences and implications for diabetes prevention interventions. Diabet. Med 2013 Apr 19;30(8):986-993. [doi: 10.1111/dme.12206]
7. Nielsen KK, Kapur A, Damm P, de Courten M, Bygbjerg IC. From screening to postpartum follow-up – the determinants and barriers for gestational diabetes mellitus (GDM) services, a systematic review. BMC Pregnancy Childbirth 2014 Jan 22;14(1). [doi: 10.1186/1471-2393-14-41]
8. Ferrara A, Hedderson MM, Brown SD, Albright CL, Ehrlich SF, Tsai A, et al. The Comparative Effectiveness of Diabetes Prevention Strategies to Reduce Postpartum Weight Retention in Women With Gestational Diabetes Mellitus: The Gestational Diabetes' Effects on Moms (GEM) Cluster Randomized Controlled Trial. Dia Care 2015 Dec 09;39(1):65-74. [doi: 10.2337/dc15-1254]
9. O'Reilly SL, Dunbar JA, Versace V, Janus E, Best JD, Carter R, et al. Mothers after Gestational Diabetes in Australia (MAGDA): A Randomised Controlled Trial of a Postnatal Diabetes Prevention Program. PLoS Med 2016 Jul 26;13(7):e1002092. [doi: 10.1371/journal.pmed.1002092]
10. Liu H, Wang L, Zhang S, Leng J, Li N, Li W, et al. One‐year weight losses in the Tianjin Gestational Diabetes Mellitus Prevention Programme: A randomized clinical trial. Diabetes Obes Metab 2018 May;20(5):1246-1255. [doi: 10.1111/dom.13225]
11. Nicklas JM, Zera CA, Seely EW, Abdul-Rahim ZS, Rudloff ND, Levkoff SE. Identifying postpartum intervention approaches to prevent type 2 diabetes in women with a history of gestational diabetes. BMC Pregnancy Childbirth 2011 Mar 24;11(1). [doi: 10.1186/1471-2393-11-23]
12. Lim K, Chi C, Chan S, Lim SL, Ang SM, Yoong JS, et al. Smart Phone APP to Restore Optimal Weight (SPAROW): protocol for a randomised controlled trial for women with recent gestational diabetes. BMC Public Health 2019 Oct 15;19(1). [doi: 10.1186/s12889-019-7691-3]
13. Nicklas J, Zera C, England L. A web-based lifestyle intervention for women with recent gestational diabetes mellitus: a randomized controlled trial. Obstet Gynecol 2014;124(3):a. [doi: 10.1097/aog.0000000000000420]
14. Villinger K, Wahl DR, Boeing H, Schupp HT, Renner B. The effectiveness of app‐based mobile interventions on nutrition behaviours and nutrition‐related health outcomes: A systematic review and meta‐analysis. Obesity Reviews 2019 Jul 28;20(10):1465-1484. [doi: 10.1111/obr.12903]

15. Henriksson P, Sandborg J, Blomberg M, Alexandrou C, Maddison R, Silfvernagel K, et al. A Smartphone App to Promote Healthy Weight Gain, Diet, and Physical Activity During Pregnancy (HealthyMoms): Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2019 Mar 01;8(3):e13011. [doi: 10.2196/13011]
16. Davis D, Davey R, Williams LT, Foureur M, Nohr E, Knight-Agarwal C, et al. Optimizing Gestational Weight Gain With the Eating4Two Smartphone App: Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2018 May 30;7(5):e146. [doi: 10.2196/resprot.9920]
17. Guo H, Zhang Y, Li P, Zhou P, Chen L, Li S. Evaluating the effects of mobile health intervention on weight management, glycemic control and pregnancy outcomes in patients with gestational diabetes mellitus. J Endocrinol Invest 2018 Nov 7;42(6):709-714. [doi: 10.1007/s40618-018-0975-0]
18. 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. Journal of Women's Health 2017 Jul;26(7):719-727. [doi: 10.1089/jwh.2016.5947]
19. Redman LM, Gilmore LA, Breaux J, Thomas DM, Elkind-Hirsch K, Stewart T, et al. Effectiveness of SmartMoms, a Novel eHealth Intervention for Management of Gestational Weight Gain: Randomized Controlled Pilot Trial. JMIR Mhealth Uhealth 2017 Sep 13;5(9):e133. [doi: 10.2196/mhealth.8228]
20. van der Pligt P, Ball K, Hesketh KD, Teychenne M, Crawford D, Morgan PJ, et al. J Hum Nutr Diet 2017 Oct 16;31(3):314-328. [doi: 10.1111/jhn.12521]
21. Chong Y, Cai S, Lin H, Soh SE, Lee Y, Leow MK, et al. Ethnic differences translate to inadequacy of high-risk screening for gestational diabetes mellitus in an Asian population: a cohort study. BMC Pregnancy Childbirth 2014 Oct 2;14(1). [doi: 10.1186/1471-2393-14-345]
22. Masao K. Measuring Asia's Mobile Transformation. Google Asia Pacific. URL: https://www.thinkwithgoogle.com/intl/en-apac/tools-resources/research-studies/measuring-asias-mobile-transformation/ [accessed 2019-03-01]
23. WHO. Diagnostic CriteriaClassification of Hyperglycaemia First Detected in Pregnancy. Organization WH. ed 2013.
24. Czajkowski S, Powell L, Adler N. From ideas to efficacy: The ORBIT model for developing behavioral treatments for chronic diseases. Health Psychol 2015;34(10):971-982. [doi: 10.1037/hea0000161]
25. Osborne RH, Elsworth GR, Whitfield K. The Health Education Impact Questionnaire (heiQ): An outcomes and evaluation measure for patient education and self-management interventions for people with chronic conditions. Patient Education and Counseling 2007 May;66(2):192-201. [doi: 10.1016/j.pec.2006.12.002]
26. Laursen DH, Christensen KB, Christensen U, Frølich A. Assessment of short and long-term outcomes of diabetes patient education using the health education impact questionnaire (HeiQ). BMC Res Notes 2017 Jun 15;10(1). [doi: 10.1186/s13104-017-2536-6]
27. Nolte S, Elsworth GR, Sinclair AJ, Osborne RH. The extent and breadth of benefits from participating in chronic disease self-management courses: A national patient-reported outcomes survey. Patient Education and Counseling 2007 Mar;65(3):351-360. [doi: 10.1016/j.pec.2006.08.016]
28. Osborne RH, Batterham R, Livingston J. The Evaluation of Chronic Disease Self-Management Support Across Settings: The International Experience of the Health Education Impact Questionnaire Quality Monitoring System. Nursing Clinics of North America 2011 Sep;46(3):255-270. [doi: 10.1016/j.cnur.2011.05.010]
29. Ware JE, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey. Medical Care 1996;34(3):220-233. [doi: 10.1097/00005650-199603000-00003]
30. Bandura A. Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.),Self-efficacy beliefs of adolescents(Vol. 5, pp. 307-337). Greenwich, CT: Information Age Publishing 2006. [doi: 10.1016/b978-0-08-044894-7.00620-5]
31. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of Attrition and Dropout in App-Based Interventions for Chronic Disease: Systematic Review and Meta-Analysis. J Med Internet Res 2020 Sep 29;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]
32. Mertens L, Braeken MAKA, Bogaerts A. Effect of Lifestyle Coaching Including Telemonitoring and Telecoaching on Gestational Weight Gain and Postnatal Weight Loss: A Systematic Review. Telemed J E Health 2018 Dec 06. [doi: 10.1089/tmj.2018.0139] [Medline: 30523742]
33. Board H. National Nutrition Survey 2010, Singapore. Health Promotion Board 2010.
34. Holmen H, Torbjørnsen A, Wahl AK, Jenum AK, Småstuen MC, Årsand E, et al. A Mobile Health Intervention for Self-Management and Lifestyle Change for Persons With Type 2 Diabetes, Part 2: One-Year Results From the Norwegian Randomized Controlled Trial RENEWING HEALTH. JMIR mHealth uHealth 2014 Dec 11;2(4):e57. [doi: 10.2196/mhealth.3882]
35. Maturi MS, Afshary P, Abedi P. Effect of physical activity intervention based on a pedometer on physical activity level and anthropometric measures after childbirth: a randomized controlled trial. BMC Pregnancy Childbirth 2011 Dec 16;11:103 [FREE Full text] [doi: 10.1186/1471-2393-11-103] [Medline: 22176722]
36. Altazan AD, Redman LM, Burton JH, Beyl RA, Cain LE, Sutton EF, et al. Mood and quality of life changes in pregnancy and postpartum and the effect of a behavioral intervention targeting excess gestational weight gain in women with overweight

and obesity: a parallel-arm randomized controlled pilot trial. BMC Pregnancy Childbirth 2019 Jan 29;19(1). [doi: 10.1186/s12884-019-2196-8]
37. Hu G, Tian H, Zhang F, Liu H, Zhang C, Zhang S, et al. Tianjin Gestational Diabetes Mellitus Prevention Program. Diabetes Research and Clinical Practice 2012 Dec;98(3):508-517. [doi: 10.1016/j.diabres.2012.09.015]
38. Dalrymple KV, Flynn AC, Relph SA, O'Keeffe M, Poston L. Lifestyle Interventions in Overweight and Obese Pregnant or Postpartum Women for Postpartum Weight Management: A Systematic Review of the Literature. Nutrients 2018 Nov 07;10(11):1704. [doi: 10.3390/nu10111704]
39. Marsaux CF, Celis-Morales C, Fallaize R, Macready AL, Kolossa S, Woolhead C, et al. Effects of a Web-Based Personalized Intervention on Physical Activity in European Adults: A Randomized Controlled Trial. J Med Internet Res 2015 Oct 14;17(10):e231. [doi: 10.2196/jmir.4660]
40. Lindström J, Peltonen M, Eriksson JG, Ilanne-Parikka P, Aunola S, Keinänen-Kiukaanniemi S, et al. Improved lifestyle and decreased diabetes risk over 13 years: long-term follow-up of the randomised Finnish Diabetes Prevention Study (DPS). Diabetologia 2012 Oct 24;56(2):284-293. [doi: 10.1007/s00125-012-2752-5]

Abbreviations

GDM: gestational diabetes mellitus
HbA_{1c}: glycated hemoglobin
heiQ: Health Education Impact Questionnaire
ITT: intention-to-treat
NUH: National University Hospital
OGTT: oral glucose tolerance test
OR: odds ratio
RCT: randomized controlled trial
SPAROW: The Smartphone App to Restore Optimum Weight
T2DM: type 2 diabetes mellitus

Edited by G Eysenbach; submitted 05.07.20; peer-reviewed by A Hidki,Dr, J Nicklas; comments to author 07.10.20; revised version received 14.12.20; accepted 09.01.21; published 16.03.21.

Please cite as:
Lim K, Chan SY, Lim SL, Tai BC, Tsai C, Wong SR, Ang SM, Yew TW, Tai ES, Yong EL
A Smartphone App to Restore Optimal Weight (SPAROW) in Women With Recent Gestational Diabetes Mellitus: Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(3):e22147
URL: https://mhealth.jmir.org/2021/3/e22147
doi:10.2196/22147
PMID:33724204

Original Paper

Intention to Adopt mHealth Apps Among Informal Caregivers: Cross-Sectional Study

Fereshteh Ghahramani[1], MPH, PhD; Jingguo Wang[2], PhD

[1]College of Computing and Digital Media, DePaul University, Chicago, IL, United States

[2]Department of Information Systems and Operations Management, University of Texas at Arlington, Arlington, TX, United States

Corresponding Author:
Fereshteh Ghahramani, MPH, PhD
College of Computing and Digital Media
DePaul University
243 South Wabash Ave
Chicago, IL, 60604
United States
Phone: 1 3123628381
Email: fereshteh.ghahramani@depaul.edu

Abstract

Background: Caregiving responsibility can change caregivers' lives; modify their emotions; and make them feel frustrated, fearful, and nervous, thereby imposing physical and mental stress. Caregiving-related mobile apps provide a platform for obtaining valuable and trusted information, connecting more easily with other caregivers, monitoring medications, and managing appointments, and assessing health requirements and conditions of care receivers. Such apps also incorporate valuable resources that address care for the caregivers. Despite the potential benefits of caregiving-related apps, only a limited number of caregivers have adopted and used them.

Objective: The aim of this study is to explore the important factors that affect caregivers' intentions to integrate related mobile apps into their routine caregiving responsibilities.

Methods: Using the protection motivation theory, we conducted a cross-sectional study among 249 participants. Purposive sampling was used to target participants who met 4 inclusion criteria: US residents, owning and using a smartphone, informal caregivers (individuals who give care to a friend or family member without payment) who provided at least 8 hours of care per week in the past year, and those currently not using any mobile app for caregiving purposes. We created a survey using Qualtrics and posted it on Amazon's Mechanical Turk website. Participants received monetary compensation after successful completion of the survey.

Results: We found that capabilities and skills of caregivers to use mobile apps, the app's effectiveness in responding to the needs of caregivers, the degree of control of caregivers over their responsibilities, and the decisions they make for their care receivers can predict their willingness to adopt caregiving-related apps. In addition, the severity of health status and vulnerability of care receivers to unexpected health changes indirectly shape their caregivers' decisions to adopt and use mobile apps for caregiving purposes.

Conclusions: This study explores the important factors that affect informal caregivers' intentions to adopt related mobile apps into their routine caregiving responsibilities. The results contribute to both mobile health adoption and the caregiving literature, and they offer significant implications for developers, health care practitioners, and policy makers.

(JMIR Mhealth Uhealth 2021;9(3):e24755) doi:10.2196/24755

KEYWORDS

mobile health; cross-sectional study; informal caregivers; mobile app; caregiving app; mobile phone

Introduction

Background

In 2017, only approximately 37 million patients were admitted to US hospitals [1]. As the number of individuals needing hospitalization has increased in recent years [2], hospitals have mainly limited their beds to acute care and left the rest of treatments in the form of outpatient settings to the patients' loved ones [3,4]. Unpaid family or informal caregivers are, therefore, the backbone of the long-term care system and offer the majority of long-term care that patients require in the United States [5]. In addition, informal caregiving is essential for sustaining adults with disabilities or chronic health conditions [6].

Caregiving responsibilities primarily include providing informational and emotional support, dealing with financial concerns, and managing medical care [4]. These responsibilities also consist of more basic support, such as assistance in eating, bathing, grocery shopping, and meal preparation [7]. Many informal caregivers experience positive feelings such as satisfaction, a sense of gratification for giving back to those who cared for them, and improved family relationships [8,9]. However, the nature and amount of such responsibilities may change caregivers' lives; modify their emotions; and make them feel frustrated, fearful, and nervous, all resulting in a decline in their quality of life [10].

Estimates of the number of informal caregivers in the United States vary widely. According to the American Association of Retired Persons, there are currently more than 40 million unpaid caregivers in the United States [11], about 7 potential informal caregivers per adult [12]. Although studies indicate that the number of informal caregivers will continue to rise, it is expected that the number of individuals who need care will be far from the number of those who offer that care in the near future (about 4 potential informal caregivers per adult) [12,13]. There are 2 main reasons for this result. First, the number of individuals who will be over 65 years old by 2050 is expected to be 2 times more than that number in 2010; the majority of those individuals will be dealing with various chronic conditions and a decline in quality of life [14,15]. Second, the traditional American family structure has experienced fundamental changes in recent years. Family sizes continue to decrease as a result of higher rates of those who have never married or are divorced or are affected by infertility and childlessness. In addition, there is a higher chance that women, who are the backbone of informal caregiving, are in the workforce [16,17].

This shrinkage in the number of potential informal caregivers per adult increases the amount of responsibility and inflicts more physical and mental stress among the existing caregivers [14]. The ubiquity of mobile technology and its applications has the potential to reduce such stress. In general, mobile apps are reasonably priced and user friendly and offer an information repository collected from various sources [18-20]. Caregiving-related apps are specifically designed to provide users with a platform to gain appropriate and trusted information, manage medication taking, improve communication with care providers and support groups, connect with counterparts, reserve transportation, and manage the health condition of care receivers in an organized manner [21].

Although there are hardly any studies in the literature that investigate the role of caregiving-related apps in reducing caregiving-related stress [22], it is very likely that such apps significantly lessen the stress caregivers face by providing a convenient platform to receive informational and emotional support [14]. Despite the considerable role of caregiving apps in reducing stress and improving the overall quality of life among caregivers and although more than 57% of American caregivers have a smartphone, only 40% of them use a caregiving-related app [23]. This raises the concern of finding solutions to increase caregivers' access to and effective use of such beneficial resources.

Some studies highlight the roles of caregivers' digital literacy and sociodemographic factors on their natural propensity to use various internet-based tools and services for caregiving purposes in general [24,25]. However, the current understanding of caregivers' intentions to use related mobile apps for their responsibilities is limited, and we could not find any published studies that directly investigated the influential factors.

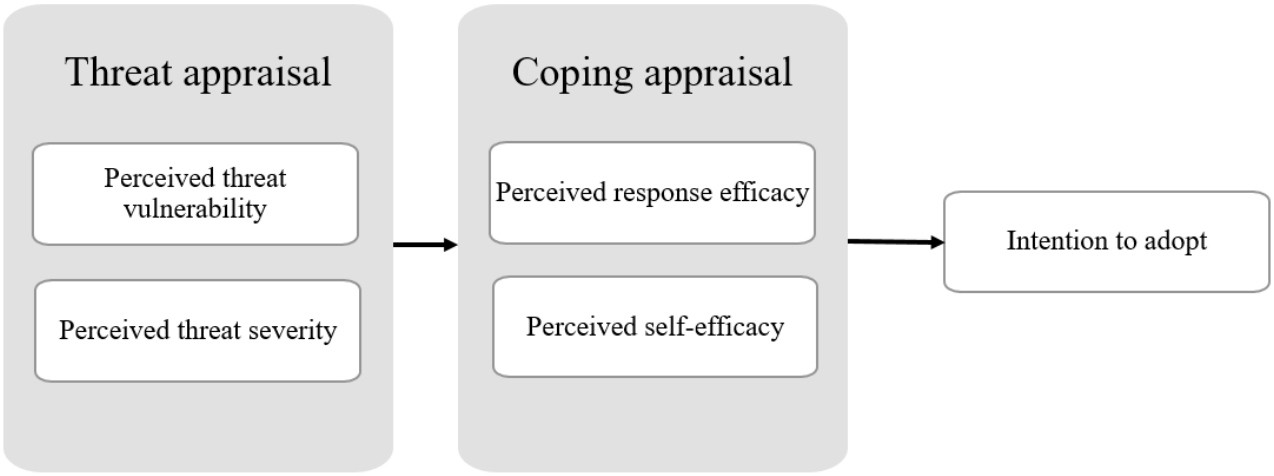
As such, our objective is to provide insights into this issue. To explore the important factors that affect their intentions to integrate mobile apps into their routine caregiving responsibilities, we designed a cross-sectional study using the protection motivation theory (PMT) perspective [26]. PMT is among the most influential explanatory theories in the literature to predict an individual's intention to adopt recommended actions (adoption of related mobile apps in this study) [27]. It has also been empirically applied to both technological and nontechnological solutions [28].

Theoretical Background

PMT suggests 2 consecutive appraisals that explain the process whereby individuals adopt a recommended action: threat and coping [29]. If an individual is exposed to a stressful or fearful situation, a personal perception of threat arises. If the threat is perceived to be appropriate and possibly harmful, coping appraisal will occur [28].

On the basis of the threatening event, individuals first assess the level of danger in 2 aspects: (1) perceived threat severity, which is the individual's assessment of the seriousness of the threat, and (2) perceived threat vulnerability, which is the individual's assessment of the likelihood of coming across the threat personally. Once the threat has been assessed, individuals assess their ability to cope with the threat and form their perceptions of the efficacy of the recommended action in 2 aspects: (1) self-efficacy, which is the individuals' assessment of their capability to perform the recommended action, and (2) response efficacy, which is the individual's assessment of the efficacy and benefits of the recommended action [26,29,30]. This efficacy assessment affects individuals' intentions to adopt the recommended action. Figure 1 illustrates the PMT model.

Figure 1. Protection motivation theory model.



Research Model and Hypotheses

Overview

Informal caregivers experience a vast amount of uncertainty and psychological distress [31] because of the difficulty in predicting their care receivers' disease and care progress [32]. Adopting and using mobile health apps can be considered a recommended action to reduce caregivers' stress. Building on PMT, we propose that, reflecting care receivers' physical and mental condition, caregivers form perceptions about the severity and vulnerability of their care receivers' health threats (threat appraisal). If they perceive a significant and harmful degree of threat, caregivers will assess both their responses and self-efficacy to cope with the situation (coping appraisal). Caregivers who expect that mobile apps can help them (high response efficacy) and have the efficacy to operate the technologies (high self-efficacy) are expected to begin using related mobile apps for caregiving purposes [28].

Moreover, as self-efficacy only reflects caregivers' perceptions of their general capability to use an app [33], we also need to consider the degree to which caregivers have control over their caregiving responsibilities and the decisions they make for their care receivers. This is called perceived self-autonomy, and previous studies have verified it as a major contributor to technology acceptance together with self-efficacy [34,35]. When caregivers feel autonomous in choosing a mobile app, they perceive it as useful and easy to adopt [34]. Thus, we propose that perceived self-autonomy is an antecedent that influences caregivers' intention to adopt and use caregiving-related apps. The proposed research model is illustrated in Figure 2. The construct definitions are presented in Table 1.

Figure 2. Research model.

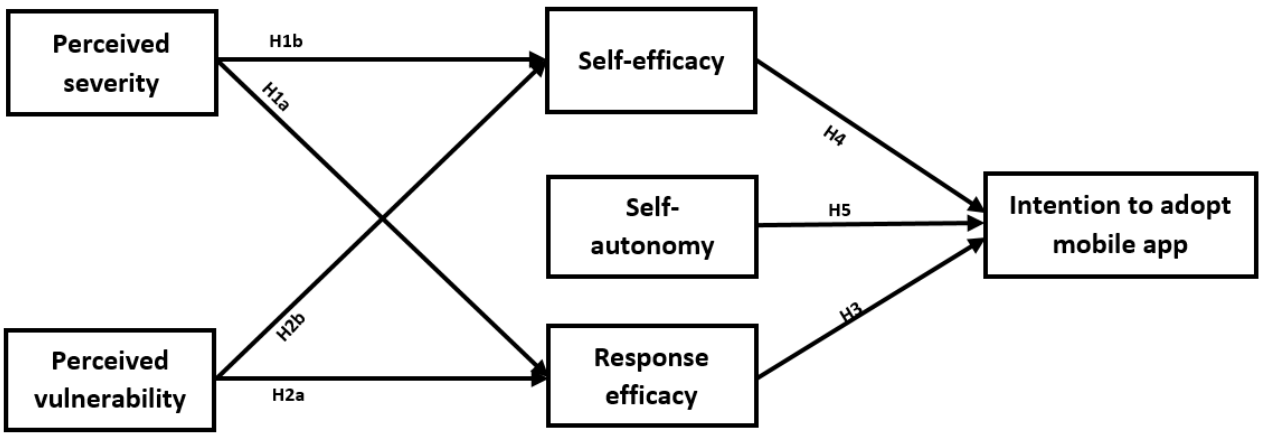


Table 1. Definitions of constructs.

Construct	Definition
Perceived threat severity	Degree to which a caregiver assesses the seriousness of their care receiver's health status
Perceived threat vulnerability	Degree to which a caregiver assesses the susceptibility of their care receiver to a sudden change in health status or an unexpected health condition
Perceived self-efficacy	Degree to which a caregiver believes that they are capable and have the necessary skills to use a caregiving-related mobile app
Perceived response efficacy	Degree to which a caregiver believes that a caregiving-related mobile app will effectively prevent threats related to their care receiver's health condition
Perceived self-autonomy	Degree to which a caregiver has control over caregiving responsibilities and the decision they make for the care receiver
Intention to adopt	Caregiver's willingness to adopt a caregiving-related app

Threat Appraisal

Depending on the physical and mental conditions of the care receiver, the caregivers assesses the seriousness of the care receiver's health status. This assessment can influence the significance of caregiving-related apps by manipulating both responses and self-efficacy perceptions of caregivers [28]. More specifically, as a caregiver feels that their care receiver deals with a more severe health situation, he or she is more willing to trust sources that can provide the support he or she is looking for to help him or her cope with the situation.

Therefore, the caregivers perceives the app as a more capable and qualified tool to effectively help him or her address the threat in a way that he or she might not have previously recognized. It can also persuade caregivers to re-examine their abilities to use the recommended app for caregiving activities. As the care receiver's health status is perceived to be more severe, the caregivers will rely more on him or her and feel more confident about using a caregiving-related app as one of the few means to deal with the severe situation. Therefore, we hypothesize the following:

Hypothesis 1a (H1a): The caregiver's perception of the care receiver's severity of health status positively influences the caregiver's perception of the mobile app's response efficacy.

Hypothesis 1b (H1b): The caregiver's perception of the care receiver's severity of health status positively influences the caregiver's perception of his or her self-efficacy in using the mobile app.

Similar to the logic of perceived severity, when caregivers notice that there is a likelihood that their care receiver will encounter a sudden and unanticipated change in their health condition, they tend to perceive the mobile app as comprehensive and effective enough to offer them the support the caregivers are looking for during this time.

Therefore, it is expected that as the threat of facing unexpected health changes becomes more probable, caregivers will perceive the mobile app as a more effective and efficient tool to help them in their caregiving responsibilities. In addition, caregivers will find themselves more capable and confident in using the mobile app as a tool to lessen the burden and the effects of the vulnerable situation. So, we proposed the following hypothesis:

Hypothesis 2a (H2a): The caregiver's perception of the care receiver's vulnerability to unexpected health changes positively influences the caregiver's perception of the mobile app's response efficacy.

Hypothesis 2b (H2b): The caregiver's perception of the care receiver's vulnerability to unexpected health changes positively influences the caregiver's perception of his or her self-efficacy in using the mobile app.

Coping Appraisal

Perceived response efficacy is a cognitive process by which caregivers develop thoughts about the effectiveness and capability of the app to address their needs while they are dealing with a threatening situation [26,29]. In other words, the caregivers' perception of the mobile app's response efficacy determines whether they choose that app to help them handle the threat or not [36]. According to PMT, a high level of response efficacy forms a positive disposition toward the recommended solution. So, we proposed the following hypothesis:

Hypothesis 3 (H3): The perceived response efficacy of the caregiving-related app has a positive effect on the caregiver's intention to adopt the app.

Even if a caregiver believes that the app is effective enough to address their needs, he or she still needs to consider their own ability to successfully install and effectively use the app [28]. Caregivers who perceive themselves as capable of using the app are more willing to adopt and use such apps to address their caregiving-related requirements [37]. Therefore, we proposed the following hypothesis:

Hypothesis 4 (H4): The caregiver's perceived self-efficacy has a positive effect on his or her intention to adopt the caregiving-related app.

Self-Autonomy

Furthermore, if caregivers have some freedom of action in their caregiving activities and have a say regarding what mobile apps they want to use, there is a higher chance that they will become intrinsically motivated or maintain the primary levels of intrinsic motivation to adopt and use the app [38]. Such a feeling of control can also enhance caregivers' positive feelings toward

using the app [39,40]. Therefore, we proposed the following hypothesis:

Hypothesis 5 (H5): The caregiver's perceived self-autonomy has a positive effect on his or her intention to adopt the caregiving-related app.

Methods

Recruitment

To test the proposed model, we recruited participants through Amazon's Mechanical Turk, which is a web-based crowdsourcing market for registered users to participate in various tasks and receive a predetermined amount of money upon successful completion [41]. Studies in the domain of information systems have confirmed the validity and reliability of the results [42,43]. Purposive sampling was used to target participants who met the following inclusion criteria [44]: (1) US residents, (2) owning and using a smartphone, (3) informal caregivers (individuals who give care to a friend or family member without payment) who provided at least 8 hours of care per week in the past year, and (4) currently not using any mobile app for caregiving purposes. We obtained approval from the Institutional Review Board for the explained approach.

Instrument Development

We adapted measures from a set of empirically validated studies in the literature and used multi-item measures to enhance the validity and reliability of the measurement. All the main constructs were reflective. For the constructs drawn from PMT, measurement items were adapted from Witte [45] and modified to fit the context of our study. Under threat appraisal, 5 items (mean 3.76, SD 0.74) and 4 items (mean 4.03, SD 0.69) were used to assess perceived vulnerability (eg, "My care-receiver is at risk for getting health threats") and perceived severity (eg, "I believe that threats to my care receiver's health are severe"), respectively. For coping appraisal, 6 items (mean 3.37, SD 0.86) assessed response efficacy (eg, "Using mobile apps is effective in monitoring my care-receiver's health condition remotely"), and 4 items (mean 3.79, SD 0.78) were used to measure self-efficacy (eg, "I feel confident using mobile health applications for my caregiving activities"). Intention to use items was adapted from a widely used scale in the literature [46] (3 items; mean 2.98, SD 0.93) and changed appropriately to fit our study (eg, "I plan to use mobile apps to manage my care-receiver's health status in the next 3 months"). Items for perceived self-autonomy were adapted from the study by Deci et al [47] (5 items; mean 3.89, SD 0.83) and altered to suit our context (eg, "I have some choice in what I want to do in my caregiving activities"). For all the aforementioned constructs, participants indicated their level of agreement with each item using a 5-point Likert-type scale ranging from *strongly disagree* (1) to s*trongly agree* (5).

Questions were randomly shown to the participants to minimize the order-effect bias. Two experts reviewed the initial questionnaire to ensure face validity. We revised the survey based on their comments and feedback. We also conducted a pilot study among master's students at a large university in the southwestern region of the United States and made appropriate changes to the survey based on the results. Previous literature, pilot tests, and numerous series of pretests allowed us to confirm the content validity of our instrument and measures [48]. The final measurement items are listed in Table 2. The Checklist for Reporting Results of Internet E-Survey is available in Multimedia Appendix 1 [49].

Table 2. Instrument items.

Construct	Items	References
Dependent variable		
Intention to use	• It is my intention to use mobile applications in caregiving activities. • I plan to use mobile apps to manage my care-receiver's health status in the next 3 months. • I am likely to learn about using mobile apps in caregiving activities.	Venkatesh et al [46]
Threat appraisal		
Vulnerability	• My care-receiver can be subjected to a sudden change in health condition. • My care-receiver is at risk for getting health threats. • It is possible that my care-receiver will contract health threats. • It is likely that my care-receiver requires an urgent care. • It is likely that my care-receiver will contract health threats.	Witte [45]
Severity	• If my care-receiver faces an unexpected health problem, it would be serious. • I believe that threats to my care-receiver's health are severe. • I believe that threats to my care-receiver's health are serious. • I believe that threats to my care-receiver's health are significant.	Witte [45]
Coping appraisal		
Response efficacy	• Mobile apps will help me manage medication for my care-receiver. • Mobile apps serve as an effective disease reference and caregiving adviser. • Mobile apps enable me to keep a log of medical information for my care-receiver. • Mobile apps work in preventing health threats due to mismanagement of medications. • Using mobile apps is effective in monitoring my care-receiver's health condition remotely (eg, heart rate, oxygen level, or other vital signs). • If I use mobile apps in my caregiving activities, my care-receiver is less likely to get health threats due to mismanagement of medications.	Witte [45]
Self-efficacy	• I feel confident using mobile health applications for my caregiving activities. • I am able to use mobile apps. • Mobile apps are easy to use. • Using mobile apps is convenient.	Witte [45]
Perceived self-autonomy	• In my caregiving activities, I can decide which mobile apps I want to use. • In my caregiving activities, I have a say regarding what mobile apps I want to use. • I feel that I will use mobile apps for caregiving purposes because I want to. • I feel a certain freedom of action in my caregiving activities. • I have some choice in what I want to do in my caregiving activities.	Deci et al [47]

Sample Characteristics

A total of 249 valid responses were collected. The average time that our responders provided care for their current care receivers and for various individuals in general was 50 and 83 months, respectively. A total of 32.5% (81/249) of the care receivers were financially dependent on their caregivers. Tables 3 and 4 summarize the key demographic variables of the respondents and their care receivers.

Table 3. Key demographics of care receivers.

Variable	Percentage, n (%)
Age (years)	
≤18	28 (11.2)
18-49	39 (15.7)
50-69	55 (22.1)
≥70	127 (51.0)
Education	
High school or general educational development	106 (42.6)
Some college or bachelor's degree	89 (35.7)
Master's degree	32 (12.9)
Professional degree	17 (6.8)
Doctoral degree	5 (2.0)
Race	
White	185 (74.3)
African American	27 (10.8)
Hispanic	22 (8.9)
Asian	11 (4.4)
Native American	2 (0.8)
Pacific Islander	2 (0.8)
Gender	
Male	102 (41.0)
Female	147 (59.0)
Relationship with care receiver	
Parents	46 (18.5)
Friend	132 (53.0)
Family friend	25 (10.0)
Spouse	28 (11.3)
Child	18 (7.2)
Reason for receiving care	
Any form of disease	78 (31.3)
Old age	77 (30.9)
Disability	59 (23.7)
Mental disorder	35 (14.1)

Table 4. Key demographics of caregivers (N=249).

Variable	Percentage, n (%)
Age (years)	
≤18	0 (0)
18-24	33 (13.3)
25-34	88 (35.3)
35-44	64 (25.7)
45-54	35 (14.1)
55-64	21 (8.4)
≥65	8 (3.2)
Education	
High school or general educational development	22 (8.8)
Some college or bachelor's degree	181 (72.7)
Master's degree	42 (16.9)
Professional degree	1 (0.4)
Doctoral degree	3 (1.2)
Race	
White	193 (77.5)
African American	24 (9.7)
Hispanic	16 (6.4)
Asian	9 (3.6)
Native American	2 (0.8)
Pacific Islander	5 (2.0)
Gender	
Male	78 (31.3)
Female	171 (68.7)
Marital status	
Single without children	72 (28.9)
Single with children	33 (13.3)
Married without children	18 (7.2)
Married with children	98 (39.4)
Life partner without children	13 (5.2)
Life partner with children	15 (6.0)
Income, US ($)	
≤20,000	47 (18.9)
20,000-40,000	72 (28.9)
40,000-60,000	61 (24.5)
60,000-80,000	41 (16.4)
≥80,000	28 (11.2)

Data Analysis

A partial least squares approach was used to test the proposed model and associated hypotheses. We used the SmartPLS software package (version 3.2.6, SmartPLS GmbH) to analyze the data [50]. To estimate the path coefficient weights and their significance, we used bootstrapping procedures with 5000 resamples.

To assess the measurement model, we conducted a confirmatory factor analysis. The results are summarized in Table 5. All of the latent constructs are modelled to be reflective. For each

latent construct, the path loadings, *t* statistics, and SE were calculated. At α=.05, all the measures' path loadings are significant with a value of more than 0.7 [51], indicating that more than 50% of the variance is shared between each construct's items [52].

Table 5. Confirmatory factor analysis results.

Construct and items	Loading	*t* test	SE	Average variance extracted	Composite reliability	Cronbach α
Intention to use				0.810	0.927	.882
Int[a]1	0.911	61.307	0.015			
Int2	0.930	107.161	0.009			
Int3	0.857	33.746	0.025			
Vulnerability				0.656	0.905	.869
Vul[b]1	0.750	10.763	0.070			
Vul2	0.834	11.179	0.075			
Vul3	0.818	9.688	0.084			
Vul4	0.767	13.247	0.058			
Vul5	0.873	17.901	0.049			
Severity				0.758	0.926	.895
Sev[c]1	0.861	18.791	0.046			
Sev2	0.835	13.664	0.061			
Sev3	0.897	23.558	0.038			
Sev4	0.888	23.897	0.037			
Response efficacy				0.681	0.927	.906
RE[d]1	0.874	63.174	0.014			
RE2	0.843	33.026	0.026			
RE3	0.842	36.862	0.023			
RE4	0.846	25.670	0.033			
RE5	0.794	20.271	0.039			
RE6	0.745	17.234	0.043			
Self-efficacy				0.636	0.874	.813
SE[e]1	0.792	30.109	0.026			
SE2	0.745	13.692	0.054			
SE3	0.787	11.989	0.066			
SE4	0.862	31.175	0.028			
Self-autonomy				0.615	0.888	.853
SA[f]1	0.837	28.530	0.029			
SA2	0.818	22.784	0.036			
SA3	0.833	43.420	0.019			
SA4	0.714	10.462	0.068			
SA5	0.708	9.575	0.074			

[a]Int: intention to use.

[b]Vul: vulnerability.

[c]Sev: severity.

[d]RE: response efficacy.

[e]SE: self-efficacy.

[f]SA: self-autonomy.

To examine convergent validity, both composite reliability and average variance extracted (AVE) were obtained [53]. All constructs met the acceptable score of 0.5 for AVE, confirming their reliability [54]. Moreover, the composite reliability values were between 0.874 and 0.927.

The square root of the AVE was also used to verify discriminant validity [54]. The results are presented in Table 6. For each latent construct, the square root of the AVE was higher than all its cross-correlations. Moreover, higher values of interconstruct correlations confirmed the greater variance among each construct's specific measures compared with other measures [55]. All things considered, we can confirm that all the indicators were valid through their constructs.

As data for all the variables were collected in a single survey, common method variance (CMV) could have an excessive effect on the results [56]. Harman's single factor test was conducted [57] to find out the extent of this effect. According to this approach, with a factor analysis, if only a single factor arises, or a factor explains the majority of variance among all the measures, we can conclude that CMV is a significant issue in the sample [56]. After conducting an exploratory factor analysis among all the items, 6 factors were obtained that explained more than 70% of the variance. All the extracted factors had an eigenvalue >1, and the highest factor accounted for only 33% of the variance. This indicates that CMV is not a serious concern in this data set.

Table 6. Matrix of latent constructs' correlations.

Construct	Mean (SD)	Construct 1	Construct 2	Construct 3	Construct 4	Construct 5	Construct 6
1. Intention to use	2.976 (0.927)	*0.900* [a]	_[b]	–	–	–	–
2. Vulnerability	3.764 (0.736)	0.609	*0.810* [a]	–	–	–	–
3. Severity	4.034 (0.691)	0.129	0.646	*0.871* [a]	–	–	–
4. Response efficacy	3.371 (0.858)	0.607	0.245	0.201	*0.825* [a]	–	–
5. Self-efficacy	3.790 (0.780)	0.479	0.111	0.208	0.652	*0.797* [a]	–
6. Self-autonomy	3.892 (0.829)	0.472	0.211	0.257	0.561	0.631	*0.784* [a]

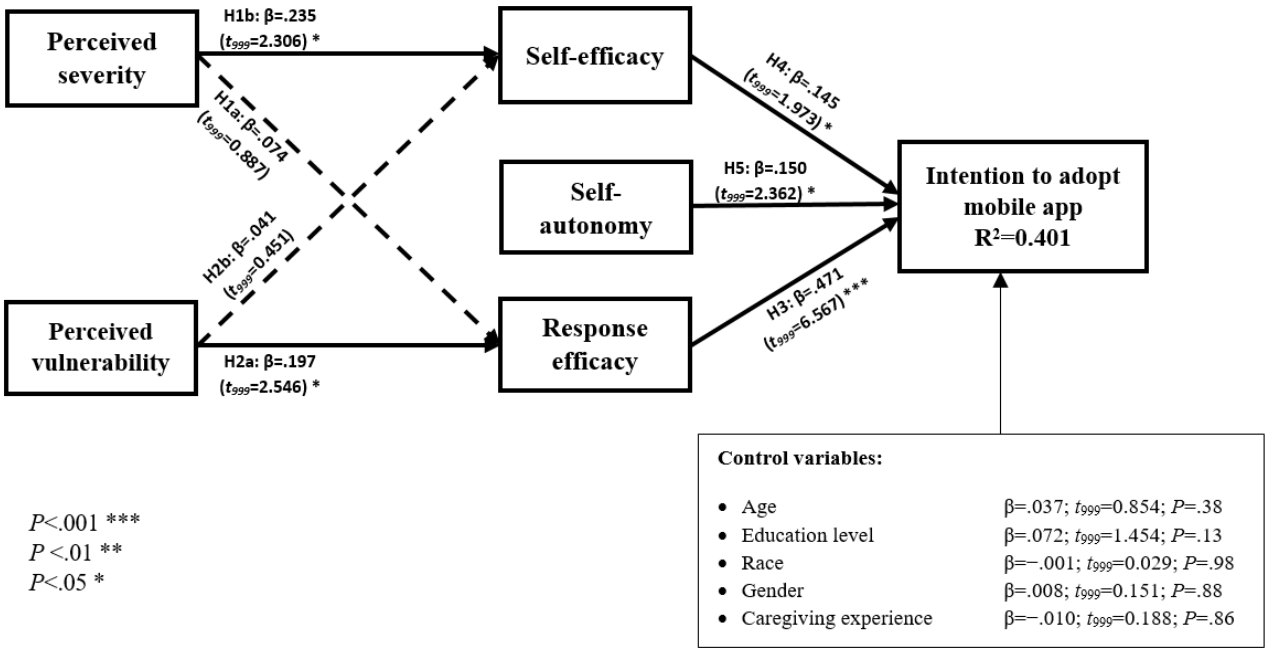
[a]Italicized values in the diagonal row are the square root of the average variance extracted.

[b]not applicable.

Results

The results of data analysis, including R^2 values, standardized path coefficients, associated *t* values, and path significance, are depicted in Figure 3. To test the significance of the path coefficients in the structural model, we used a bootstrapping approach with 1000 resamples.

Figure 3. Partial least squares results.



The R^2 value of intention to adopt a caregiving mobile app is 0.401, indicating that more than 40% of the variance in intention to adopt can be explained by response efficacy, self-efficacy, and self-autonomy. As this amount of variance is more than

10%, we can claim that the proposed model is valid and acceptable [58].

The results indicate that response efficacy (β=.471; t_{999}=6.567; *P*<.001), self-efficacy (β=.145; t_{999}=1.973; *P*=.049), and self-autonomy (β=.150; t_{999}=2.362; *P*=.03) had significant effects on the intention to adopt, supporting H3, H4, and H5. Moreover, perceived vulnerability and perceived severity had significant effects on response efficacy (β=.197; t_{999}=2.546; *P*=.008) and self-efficacy (β=.235; t_{999}=2.306; *P*=.01), respectively, supporting H2a and H1b. Perceived vulnerability and perceived severity, however, do not significantly affect self-efficacy (β=.041; t_{999}=0.451; *P*=.63) and response efficacy (β=.074; t_{999}=0.887; *P*=.35), respectively. Therefore, we do not have sufficient evidence to support H2b and H1a. The results are summarized in Table 7.

To account for individual differences in the proposed model, we controlled for the effects of the caregivers' age, education level, race, gender [46,58], and caregiving experience. None of these variables had a significant effect on the intention to adopt a caregiving mobile app.

Table 7. Summary of results.

Hypothesis	Result
H1a. The caregiver's perception of the care receiver's severity of health status positively influences the caregiver's perception of the mobile app's response efficacy.	Not supported
H1b. The caregiver's perception of the care receiver's severity of health status positively influences the caregiver's perception of his or her self-efficacy to use the mobile app.	Supported
H2a. The caregiver's perception of the care receiver's vulnerability to unexpected health changes positively influences the caregiver's perception of the mobile app's response efficacy.	Supported
H2b. The caregiver's perception of the care receiver's vulnerability to unexpected health changes positively influences the caregiver's perception of his or her self-efficacy to use the mobile app.	Not supported
H3. The perceived response efficacy of the caregiving-related app has a positive effect on the caregiver's intention to adopt the app.	Supported
H4. The caregiver's perceived self-efficacy has a positive effect on his or her intention to adopt the caregiving-related app.	Supported
H5. The caregiver's perceived self-autonomy has a positive effect on his or her intention to adopt the caregiving-related app.	Supported

Discussion

Principal Findings

The primary goal of this study is to investigate the factors that affect caregivers' intentions to adopt caregiving-related mobile apps based on a model that contextualizes PMT. The results indicate support for the proposed model with decent explanatory power.

As the analyses illustrate, caregivers' capabilities and skills to use mobile apps and the app's effectiveness in responding to caregivers' needs can predict their willingness to adopt related apps. In addition, our results indicate that the degree of control of caregivers over their responsibilities and the decisions they make for their care receivers can also increase the likelihood of adopting such mobile apps. Interestingly, the app's effectiveness in responding to caregivers' needs had the strongest effect on their intention to adopt such apps.

We also found that as the care receiver's health status is perceived to be more severe, caregivers will count more on their capability to use a caregiving-related app. In addition, as the threat of facing unexpected health changes becomes more likely, caregivers will perceive the mobile app as a more efficient tool to help them with their responsibilities. These findings are consistent with previous studies on PMT [28].

However, we did not find enough evidence to support the effect of the care receiver's severity of health status on the caregiver's perception of the mobile app's effectiveness. In addition, the results indicate that the effect of the care receiver's vulnerability to unexpected health changes on the caregiver's perception of their self-efficacy in using the mobile app is not significant. Although these results are not congruent with PMT, there are various other studies that confirm that individuals, in most cases, believe that threats either happen only to others or influence other individuals more than themselves [59]. This belief helps them maintain a sense of invulnerability and explains why we did not find evidence to support 2 of the hypotheses.

Implications

As mentioned earlier, caregiving apps play a considerable role in reducing stress and improving the overall quality of life of informal caregivers. Therefore, it is important to devote time, money, and effort to develop and promote caregiving-related apps to enhance both caregivers' and care receivers' well-being. In this regard, the results of this study provide several practical implications for developers, health care practitioners, and policy makers.

Developers

Currently, more than 318,000 mobile health apps are available to consumers in the top app stores globally, and more than 200 apps are being added every day [60]. However, only a small number of these apps are designed specifically to help caregivers with the challenges they face because of their responsibilities (excluding apps for professional caregiving organizations or those that help in locating such services and organizations) [14].

Developing more caregiving-related apps gives caregivers the opportunity to have several options to choose from and enhances their sense of control over a stressful situation.

Moreover, caregiving-related apps must offer a comprehensive set of features that caregivers typically look for. Although previous studies have confirmed the significant and positive effects of caregivers' quality of life on the patients' quality of life [61], only 20% of current caregiving-related apps have a collective set of features required by caregivers at different stages of their responsibilities [14]. According to our results, response efficacy has the strongest effect on behavioral intention. Including a broad range of features that address care for the care receiver (ie, appointment and medication management and reliable information about the disease or specific situation) can increase the chance of the caregivers' adoption of those apps.

In addition to the resources and features on how to manage their care receivers' condition, incorporating features that address care for the caregivers can help them find a way to meet their needs in one place. This can increase the effectiveness of mobile apps [62] and possibly increase the likelihood of adopting these apps. For instance, apps may contain features to (1) learn and assess emotions such as journaling as a tool to keep track of moods and mood shifts; (2) manage and reduce personal stress through quick meditative activities and breathing exercises that fit into busy routines or through the use of in-app coloring books; (3) receive emotional and social support such as app-based chats and support groups [14]; and (4) receive informational support such as podcasts or discussions on caregivers' self-care, including suggestions, resources, and inspiring words from others in the community of caregivers.

Finally, based on our results, caregiving-related apps should be designed in a user-friendly, straightforward way to help caregivers locate services more easily, considering the time constraints of caregivers. Moreover, nearly 55% of caregivers are aged 50 years or more [63]. This makes it more important to design such apps in a simple and easy-to-operate manner to ensure that both younger and older adults can equally benefit from those apps.

Health Care Practitioners

Lack of awareness about the appropriate options or possible benefits, feeling overwhelmed by the available choices, and shortness of time to conduct appropriate research are among the biggest barriers of caregivers' adoption of technology [64]. By educating physicians, nurses, social workers, and personal care aides on the available caregiving apps and asking them to spread the word by suggesting approved ones, there is a higher chance that those apps will find their way into caregivers' routines.

Moreover, based on our results, awareness of the care receiver's health condition and how it may progress can affect the perceived effectiveness and efficiency of the suggested app among caregivers. Rather than only suggesting the app, it would be better to suggest why and how (based on the severity of the care receiver's health status and vulnerability to unexpected health changes) the suggested app can take some weight off of caregivers' shoulders. This indirectly increases the chances of adopting such apps.

Flyers and signage are also great tools to target caregivers and to let them know the names and features of available caregiving-related apps in the market. Information about the variety of services offered by the apps, assurance of consistency with federal or state security and privacy acts, and instructions on ways to find and download the app (ie, relevant screenshots and licensed QR codes) are examples of the essential information that should be included in such flyers.

Policy Makers

Besides spreading the word to the worlds of developers and health practitioners, implementing various incentives might be a good motivator to increase adoption rates among caregivers. Insurance providers and policy makers can take into account policies to promote the adoption and use of caregiving-related apps [65]. Covering the cost of caregiving-related app purchases by insurance companies, reducing the price of apps or in-app purchases, raising research funds to develop more apps in this category, setting standards to require a minimum level of quality for those apps, and supporting young and motivated developers who are interested in this area are only a few examples of systematic changes that increase the likelihood of adopting and using caregiving-related apps [66].

Limitations and Future Research

This study has some limitations. First, the survey was limited to US caregivers, raising the external validity issue. Future research is required to generalize the results and offer an inclusive perspective of caregivers' adoption of related mobile apps in the context of other countries. Second, the collected information and derived results were based on self-reported data. This can increase the likelihood of some biases such as social desirability. Further studies may use a more inclusive data collection technique to measure adoption more rigorously. Finally, there is still some unexplained variance in the intention to adopt mobile apps. Although including more constructs may have improved the explanatory factor of the proposed model, we intended to maintain a parsimonious extension of PMT [28]. Additional studies are required to investigate other effective factors.

Conclusions

Grounded on PMT, this study investigates the factors that affect the intentions of informal caregivers to adopt related mobile apps for their routine caregiving responsibilities. The results of the survey of 249 US-based informal caregivers indicated that caregivers' degree of control over their responsibilities and choices, their perception of the app's effectiveness in responding to their needs, and their capability to use mobile apps play a positive role in their willingness to adopt caregiving-related apps. Furthermore, care receivers' vulnerability to unpredicted health changes and the severity of their health status indirectly affect their caregivers' intentions to adopt such mobile apps. These findings offer important contributions to the field and have significant implications for developers, health care practitioners, and policy makers.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The checklist for reporting results of our internet e-survey (CHERRIES).
[DOCX File , 16 KB - mhealth_v9i3e24755_app1.docx]

References

1. Fast facts on US hospitals. American Hospital Association. 2020. URL: https://www.aha.org/statistics/fast-facts-us-hospitals [accessed 2021-03-08]
2. Total number of hospital admissions in the U.S. from 1946 to 2018. 2020. URL: https://www.statista.com/statistics/459718/total-hospital-admission-number-in-the-us/ [accessed 2020-04-28]
3. Kane RL, Priester R, Totten AM. Meeting the challenge of chronic illness. Baltimore: The Johns Hopkins University Press; 2006.
4. American Cancer Society. 2019. URL: https://www.cancer.org/research/cancer-facts-statistics/survivor-facts-figures.html [accessed 2021-02-24]
5. Robison J, Fortinsky R, Kleppinger A, Shugrue N, Porter M. A broader view of family caregiving: effects of caregiving and caregiver conditions on depressive symptoms, health, work, and social isolation. J Gerontol B Psychol Sci Soc Sci 2009 Nov 24;64(6):788-798. [doi: 10.1093/geronb/gbp015] [Medline: 19318470]
6. Caregiving. Family Caregiver Alliance. 2009. URL: https://www.caregiver.org/caregiving [accessed 2021-02-24]
7. Flinn B. Millennials: the emerging generation of family caregivers. Innovation in Aging 2018:1-240. [doi: 10.1093/geroni/igy023.896]
8. Abdollahpour I, Nedjat S, Salimi Y. Positive aspects of caregiving and caregiver burden: a study of caregivers of patients with dementia. J Geriatr Psychiatry Neurol 2018 Jan 29;31(1):34-38. [doi: 10.1177/0891988717743590] [Medline: 29187025]
9. Long-term care in America: expectations and preferences for care and caregiving. NORC. 2016. URL: https://www.longtermcarepoll.org/project/long-term-care-in-america-expectations-and-preferences-for-care-and-caregiving/ [accessed 2021-03-08]
10. van Ryn M, Sanders S, Kahn K, van Houtven C, Griffin JM, Martin M, et al. Objective burden, resources, and other stressors among informal cancer caregivers: a hidden quality issue? Psychooncology 2011 Jan 22;20(1):44-52 [FREE Full text] [doi: 10.1002/pon.1703] [Medline: 20201115]
11. Jenkins JA. Millennials and caregiving? Yes, it's a thing. American Association of Retired Persons. 2018. URL: https://www.aarp.org/caregiving/home-care/info-2018/millennial-caregivers-work-life.html [accessed 2021-03-08]
12. Caregiving for family and friends — a public health issue. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/aging/agingdata/docs/caregiver-brief-508.pdf [accessed 2021-02-24]
13. Raskin R. 40 million caregivers and not growing strong. HuffPost. 2017. URL: https://www.huffpost.com/entry/40-million-caregivers-and_b_9019978 [accessed 2021-02-24]
14. Grossman MR, Zak DK, Zelinski EM. Mobile apps for caregivers of older adults: quantitative content analysis. JMIR Mhealth Uhealth 2018 Jul 30;6(7):e162 [FREE Full text] [doi: 10.2196/mhealth.9345] [Medline: 30061093]
15. The state of aging and health in America 2013. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/aging/pdf/state-aging-health-in-america-2013.pdf [accessed 2021-02-24]
16. Redfoot D, Feinberg L, Houser A. The aging of the baby boom and the growing care gap: a look at future declines in the availability of family caregivers. AARP Public Policy Institute. URL: https://www.aarp.org/content/dam/aarp/research/public_policy_institute/ltc/2013/baby-boom-and-the-growing-care-gap-insight-AARP-ppi-ltc.pdf [accessed 2021-02-24]
17. Committee on Family Caregiving for Older Adults, Board on Health Care Services; Health and Medicine Division, National Academies of Sciences, Engineering, and Medicine. Families caring for an aging America. National Academies Press 2016. [doi: 10.17226/23606] [Medline: 27905704]
18. Frisbee KL. Variations in the use of mHealth tools: the VA mobile health study. JMIR Mhealth Uhealth 2016 Jul 19;4(3):e89 [FREE Full text] [doi: 10.2196/mhealth.3726] [Medline: 27436165]
19. Portz JD, Elsbernd K, Plys E, Ford KL, Zhang X, Gore MO, et al. Elements of social convoy theory in mobile health for palliative care: scoping review. JMIR Mhealth Uhealth 2020 Jan 06;8(1):- [FREE Full text] [doi: 10.2196/16060] [Medline: 31904581]
20. Dayer L, Heldenbrand S, Anderson P, Gubbins PO, Martin BC. Smartphone medication adherence apps: potential benefits to patients and providers. J Am Pharm Assoc 2013 Mar;53(2):172-181. [doi: 10.1331/japha.2013.12202]
21. Saltzman M. These apps for caregivers can help you get organized, find support. American Association of Retired Persons. URL: https://www.aarp.org/home-family/personal-technology/info-2019/top-caregiving-apps.html [accessed 2020-03-06]

22. Wittenberg E, Xu J, Goldsmith J, Mendoza Y. Caregiver communication about cancer: development of a mhealth resource to support family caregiver communication burden. Psychooncology 2019 Feb 11;28(2):365-371 [FREE Full text] [doi: 10.1002/pon.4950] [Medline: 30471236]
23. Wang H, Hanich K, Jiang Y, Mitchel D, Li K. Independent living needs: helping aging parents thrive. 2016. URL: https://www.parksassociates.com/marketfocus/independent-living-360vu-march2016 [accessed 2021-02-24]
24. Chiu TML, Eysenbach G. Theorizing the health service usage behavior of family caregivers: a qualitative study of an internet-based intervention. Int J Med Inform 2011 Nov;80(11):754-764. [doi: 10.1016/j.ijmedinf.2011.08.010]
25. Shaffer KM, Chow PI, Cohn WF, Ingersoll KS, Ritterband LM. Informal caregivers' use of internet-based health resources: an analysis of the health information national trends survey. JMIR Aging 2018 Dec 18;1(2):- [FREE Full text] [doi: 10.2196/11051] [Medline: 31518244]
26. Rogers RW. A protection motivation theory of fear appeals and attitude change1. J Psychol 1975 Sep;91(1):93-114. [doi: 10.1080/00223980.1975.9915803] [Medline: 28136248]
27. Anderson CL, Agarwal R. Practicing safe computing: a multimethod empirical examination of home computer user security behavioral intentions. MIS Q 2010;34(3):613. [doi: 10.2307/25750694]
28. Johnston AC, Warkentin M. Fear appeals and information security behaviors: an empirical study. MIS Q 2010;34(3):549. [doi: 10.2307/25750691]
29. Witte K. Putting the fear back into fear appeals: the extended parallel process model. Commun Monogr 1992 Dec;59(4):329-349. [doi: 10.1080/03637759209376276]
30. Witte K. Fear control and danger control: a test of the Extended Parallel Process Model (EPPM). Commun Monogr 2009 Jun 02;61(2):113-134. [doi: 10.1080/03637759409376328]
31. Schulz R, O'Brien AT, Bookwala J, Fleissner K. Psychiatric and physical morbidity effects of dementia caregiving: prevalence, correlates, and causes. Gerontologist 1995 Dec 01;35(6):771-791. [doi: 10.1093/geront/35.6.771] [Medline: 8557205]
32. Poulshock SW, Deimling GT. Families caring for elders in residence: issues in the measurement of burden. J Gerontol 1984 Mar 01;39(2):230-239. [doi: 10.1093/geronj/39.2.230] [Medline: 6699382]
33. Agarwal R, Karahanna E. Time flies when you're having fun: cognitive absorption and beliefs about information technology usage. MIS Q 2000 Dec;24(4):665. [doi: 10.2307/3250951]
34. Cheon J, Lee S, Crooks SM, Song J. An investigation of mobile learning readiness in higher education based on the theory of planned behavior. Comput Educ 2012 Nov;59(3):1054-1064. [doi: 10.1016/j.compedu.2012.04.015]
35. Liaw SS, Huang H, Chen G. An activity-theoretical approach to investigate learners' factors toward e-learning systems. Comput Hum Behav 2007 Jul;23(4):1906-1920. [doi: 10.1016/j.chb.2006.02.002]
36. Rogers RW, Rogers R, Cacioppo BL, Petty LL. Cognitive and psychological processes in fear appeals and attitude change: a revised theory of protection motivation. Soc Psychol. 1983. URL: https://www.scienceopen.com/document?vid=7533a33e-b921-43d9-9b18-6b3f5c52f5f7 [accessed 2021-02-24]
37. Maddux JE, Rogers RW. Protection motivation and self-efficacy: a revised theory of fear appeals and attitude change. J Exp Soc Psychol 1983 Sep;19(5):469-479. [doi: 10.1016/0022-1031(83)90023-9]
38. Roca JC, Gagné M. Understanding e-learning continuance intention in the workplace: a self-determination theory perspective. Comput Hum Behav 2008 Jul;24(4):1585-1604. [doi: 10.1016/j.chb.2007.06.001]
39. Joussemet M, Koestner R, Lekes N, Houlfort N. Introducing uninteresting tasks to children: a comparison of the effects of rewards and autonomy support. J Pers 2004 Feb;72(1):139-166. [doi: 10.1111/j.0022-3506.2004.00259.x] [Medline: 14686887]
40. Flavián C, Guinalíu M. Consumer trust, perceived security and privacy policy. Industr Mngmnt & Data Systems 2006 Jun;106(5):601-620. [doi: 10.1108/02635570610666403]
41. Bohannon J. Human subject research. Social science for pennies. Science 2011 Oct 21;334(6054):307. [doi: 10.1126/science.334.6054.307] [Medline: 22021834]
42. Steelman ZR, Hammer BI, Limayem M. Data collection in the digital age: innovative alternatives to student samples. MIS Q 2014 Feb 2;38(2):355-378. [doi: 10.25300/MISQ/2014/38.2.02]
43. Lowry PB, D'Arcy J, Hammer B, Moody GD. "Cargo Cult" science in traditional organization and information systems survey research: a case for using nontraditional methods of data collection, including Mechanical Turk and online panels. J Strat Info Sys 2016 Oct;25(3):232-240. [doi: 10.1016/j.jsis.2016.06.002]
44. Patton MQ. Qualitative research & evaluation methods. Thousand Oaks, CA: SAGE Publications; 2002. URL: https://in.sagepub.com/en-in/sas/qualitative-research-evaluation-methods/book232962 [accessed 2021-02-24]
45. Witte K, Cameron KA, McKeon JK, Berkowitz JM. Predicting risk behaviors: development and validation of a diagnostic scale. J Health Commun 1996 Oct 29;1(4):317-341. [doi: 10.1080/108107396127988] [Medline: 10947367]
46. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]
47. Deci EL, Ryan RM, Gagné M, Leone DR, Usunov J, Kornazheva BP. Need satisfaction, motivation, and well-being in the work organizations of a former eastern bloc country: a cross-cultural study of self-determination. Pers Soc Psychol Bull 2016 Jul 02;27(8):930-942. [doi: 10.1177/0146167201278002]

48. Straub D, Boudreau MC, Gefen D. Validation guidelines for IS positivist research. Commu Asso Info Sys 2004;13. [doi: 10.17705/1cais.01324]
49. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet e-surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
50. Ringle C, Wende S, Becker J. SmartPLS3. Boenningstedt: SmartPLS GmbH. 2015. URL: https://www.smartpls.com/ [accessed 2021-02-24]
51. Chin W, Marcolin B. The holistic approach to construct validation in is research: examples of the interplay between theory and measurement 1. Mathematics 1995:34 [FREE Full text] [doi: 10.4135/9781452224749.n8]
52. Chin WW. Commentary: issues opinion on structural equation modeling. MIS Q. vol. 22, no. 1, pp. vii-xvi. 1998. URL: https://www.scirp.org/(S(351jmbntvnsjt1aadkposzje))/reference/ReferencesPapers.aspx?ReferenceID=2126824 [accessed 2021-02-24]
53. Hair Jr. JF. Multivariate data analysis with readings, 4th ed. NJ: Prentice Hall: Englewood Cliffs; 1995.
54. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Mark Res 2018 Nov 28;18(1):39-50. [doi: 10.1177/002224378101800104]
55. Gefen D, Straub D, Boudreau M. Structural equation modeling and regression: guidelines for research practice. Commu Asso Info Sys 2000;4. [doi: 10.17705/1cais.00407]
56. Podsakoff PM, MacKenzie SB, Lee J, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol 2003 Oct;88(5):879-903. [doi: 10.1037/0021-9010.88.5.879] [Medline: 14516251]
57. Podsakoff PM, Organ DW. Self-reports in organizational research: problems and prospects. J Manage 2016 Sep 04;12(4):531-544. [doi: 10.1177/014920638601200408]
58. Falk RF, Miller NB. A primer for soft modeling. United States: University of Akron Press; 1992.
59. Loch KD, Carr HH, Warkentin ME. Threats to information systems: today's reality, yesterday's understanding. MIS Q 1992 Jun;16(2):173. [doi: 10.2307/249574]
60. Aitken M, Clancy B, Nass D. The growing value of digital health in the United Kingdom. Evidence and Impact on Human Health and the Healthcare System Institute Report. 2017. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports/the-growing-value-of-digital-health-in-the-united-kingdom [accessed 2021-02-24]
61. Alejandra CU, Alessandrini M, Urzúa A, Zendjidjian X, Boyer L, Williams DR. Caregiver's quality of life and its positive impact on symptomatology and quality of life of patients with schizophrenia. Health Qual Life Outcomes 2017 Apr 19;15(1):76 [FREE Full text] [doi: 10.1186/s12955-017-0652-6] [Medline: 28424076]
62. Montgomery RV, Kwak J, Kosloski KD. Theories guiding support services for family caregivers. In: Handbook of Theories of Aging. New York: Springer Publishing Company; 2016:443-462.
63. Caregiving in the U.S. 2020 Report. American Association of Retired Persons. 2020. URL: https://www.aarp.org/content/dam/aarp/ppi/2020/05/full-report-caregiving-in-the-united-states.doi.10.26419-2Fppi.00103.001.pdf [accessed 2021-02-24]
64. Seegert L. Study: caregivers want tech support, but adoption rates low. Association of Health Care Journalists. 2016. URL: https://healthjournalism.org/blog/2016/06/study-caregivers-want-tech-support-but-adoption-rates-low/ [accessed 2020-09-15]
65. Kim H, Powell MP, Bhuyan SS. Seeking medical information using mobile apps and the internet: are family caregivers different from the general public? J Med Syst 2017 Mar 18;41(3):38. [doi: 10.1007/s10916-017-0684-9] [Medline: 28101781]
66. Ghahramani F, Wang J. Impact of smartphones on quality of life: a health information behavior perspective. Inf Syst Front 2019 Jun 03;22(6):1275-1290. [doi: 10.1007/s10796-019-09931-z]

Abbreviations

AVE: average variance extracted
CMV: common method variance
PMT: protection motivation theory

Edited by L Buis; submitted 03.10.20; peer-reviewed by R Mousavi, W Tang; comments to author 27.10.20; revised version received 14.12.20; accepted 12.02.21; published 17.03.21.

Please cite as:
Ghahramani F, Wang J
Intention to Adopt mHealth Apps Among Informal Caregivers: Cross-Sectional Study
JMIR Mhealth Uhealth 2021;9(3):e24755
URL: https://mhealth.jmir.org/2021/3/e24755
doi:10.2196/24755
PMID:33729166

Original Paper

Interpretable Conditional Recurrent Neural Network for Weight Change Prediction: Algorithm Development and Validation Study

Ho Heon Kim[1], MS; Youngin Kim[2], MD; Yu Rang Park[1], PhD

[1]Department of Biomedical Systems Informatics, College of Medicine, Yonsei University, Seoul, Republic of Korea

[2]Noom Inc, New York, NY, United States

Corresponding Author:
Yu Rang Park, PhD
Department of Biomedical Systems Informatics
College of Medicine
Yonsei University
50-1 Yonsei-ro
Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 0222282493
Email: yurangpark@yuhs.ac

Abstract

Background: In recent years, mobile-based interventions have received more attention as an alternative to on-site obesity management. Despite increased mobile interventions for obesity, there are lost opportunities to achieve better outcomes due to the lack of a predictive model using current existing longitudinal and cross-sectional health data. Noom (Noom Inc) is a mobile app that provides various lifestyle-related logs including food logging, exercise logging, and weight logging.

Objective: The aim of this study was to develop a weight change predictive model using an interpretable artificial intelligence algorithm for mobile-based interventions and to explore contributing factors to weight loss.

Methods: Lifelog mobile app (Noom) user data of individuals who used the weight loss program for 16 weeks in the United States were used to develop an interpretable recurrent neural network algorithm for weight prediction that considers both time-variant and time-fixed variables. From a total of 93,696 users in the coaching program, we excluded users who did not take part in the 16-week weight loss program or who were not overweight or obese or had not entered weight or meal records for the entire 16-week program. This interpretable model was trained and validated with 5-fold cross-validation (training set: 70%; testing: 30%) using the lifelog data. Mean absolute percentage error between actual weight loss and predicted weight was used to measure model performance. To better understand the behavior factors contributing to weight loss or gain, we calculated contribution coefficients in test sets.

Results: A total of 17,867 users' data were included in the analysis. The overall mean absolute percentage error of the model was 3.50%, and the error of the model declined from 3.78% to 3.45% by the end of the program. The time-level attention weighting was shown to be equally distributed at 0.0625 each week, but this gradually decreased (from 0.0626 to 0.0624) as it approached 16 weeks. Factors such as usage pattern, weight input frequency, meal input adherence, exercise, and sharp decreases in weight trajectories had negative contribution coefficients of –0.021, –0.032, –0.015, and –0.066, respectively. For time-fixed variables, being male had a contribution coefficient of –0.091.

Conclusions: An interpretable algorithm, with both time-variant and time-fixed data, was used to precisely predict weight loss while preserving model transparency. This week-to-week prediction model is expected to improve weight loss and provide a global explanation of contributing factors, leading to better outcomes.

(JMIR Mhealth Uhealth 2021;9(3):e22183) doi:10.2196/22183

KEYWORDS

explainable AI; interpretable AI; mHealth; obesity; behavior modification; artificial intelligence; development; validation; weight; intervention

Introduction

In the last 30 years, the prevalence of global obesity has increased [1]. The increasing prevalence has been observed in all countries, whether low-and-middle or high-income [2]. In the United States, the prevalence has increased in children as well as adults [3]. Because large populations with obesity are at risk of comorbid conditions, such as cardiovascular disease, gastrointestinal disorders, type 2 diabetes, and psychological issues that may increase the risk of mortality and affect their daily lives [4], this public health problem poses burdens from both an economic and health perspective [5].

As an alternative to conventional obesity management, which requires visiting medical institutions and medical staff, more attention has been given to smartphone-based interventions due to their accessibility (providing low-cost service to large populations anytime and anywhere) [6,7]. With the increase in obesity management using smartphones, a number of apps aimed toward obesity are being developed [8]. Many existing apps have collected patient-generated health data over time, including passive or active data. These data may help predict outcomes because they are from a heterogeneous population and can capture high-resolution information outside clinical settings [9-11]; however, the use of these data for weight loss prediction has not been sufficiently researched using an obesity management app.

To forecast medical outcomes such as weight loss, artificial intelligence (AI) has been adopted in the medical field [12-14]. Specifically, in AI modeling, recurrent neural networks (RNNs) have been used with longitudinal data [12,15,16]. Recently, to make RNNs transparent, more attention has been focused on explainable AI for critical safety applications in the medical field [17]. However, although time-fixed data such as demographic data, depend on each patient, and it is important in the medical field, no attempt has been made to use both serial and time-fixed data while preserving model transparency [16]. Within the same context that doctors apply while taking patient histories and monitoring consecutive laboratory examinations, it is necessary to use both types of medical data. From 2 types of data, explainable AI can act as an interpreter for the results or findings and translate them into a format that can easily be understood while maintaining performance for timely feedback in obesity management [18,19]. Furthermore, the explanation from this model is a prerequisite for new insight into patterns in obesity management.

In this study, we aimed to (1) develop a predictive model by utilizing both time-variant and time-fixed data for weight loss prediction in a mobile-based intervention using interpretable AI and (2) explore the factors affecting weight loss or gain based on app usage.

Methods

Mobile App

Noom (Noom Inc) is a mobile app that provides various lifestyle-related logs and is available for Android and Apple devices. The app provides interventions based on (1) food logging, (2) exercise logging, (3) weight logging, (4) in-app group activities, (5) article reading, and (6) communication with coaches through messaging [20]. Once users sign up and install the app, they are required to record their initial status information such as weight, BMI, and target weight. The coach encourages users to record their food intake, daily exercise, and weekly weight [20]. Furthermore, passive data, such as the number of steps taken, are automatically collected and saved in order to track users' behaviors.

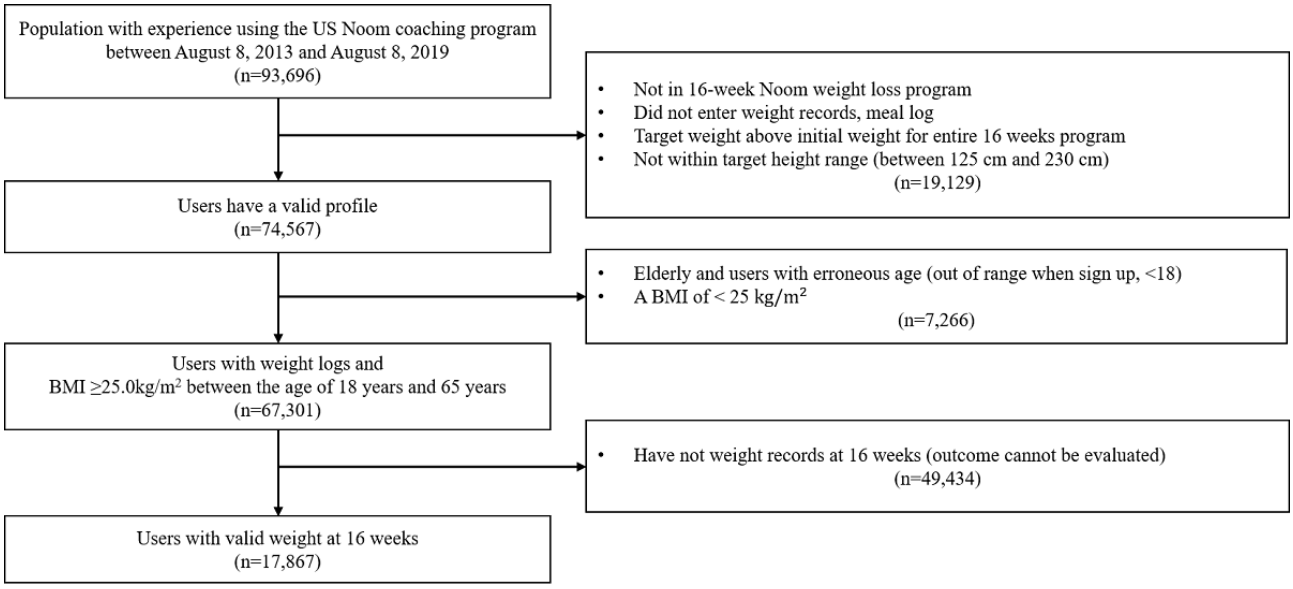
Study Design and Participants

We obtained anonymous and deidentified log data of Noom users in the United States from August 8, 2013 to August 8, 2019 for a retrospective study. This study adhered to the TRIPOD [21] statement on reporting predictive models.

From a total of 93,696 users in the coaching program, we excluded users who did not take part in the 16-week weight loss program. From this group, we ruled out users who had not entered weight or meal records for the entire 16-week program. Although the coach encouraged users to enter their meal and weight log daily, those with incomplete records were considered users who did not successfully receive the intervention; moreover, we could not evaluate their outcomes in the model for weight loss at 16 weeks. We included users who had a target weight that was below the initial weight to exclude users with the purpose of weight gain. In addition, we excluded those whose height did not fall between 125 cm and 230 cm (n=19,129).

In users with a valid profile, we included only adult users (range: 18 to 65 years old) and excluded older adults over 65 years old and users under 18 years old. Furthermore, only overweight and obese users (BMI ≥25 kg/m^2 based on the definitions provided by the Center for Disease Control and Prevention [22]) were included in our prediction model to target users with a need for weight loss (n=67,301). Finally, we identified 17,867 eligible users with 16 weeks of weight records for analysis to evaluate the outcome, presented as predicted weight (Figure 1).

Figure 1. Selection flow for eligible users.

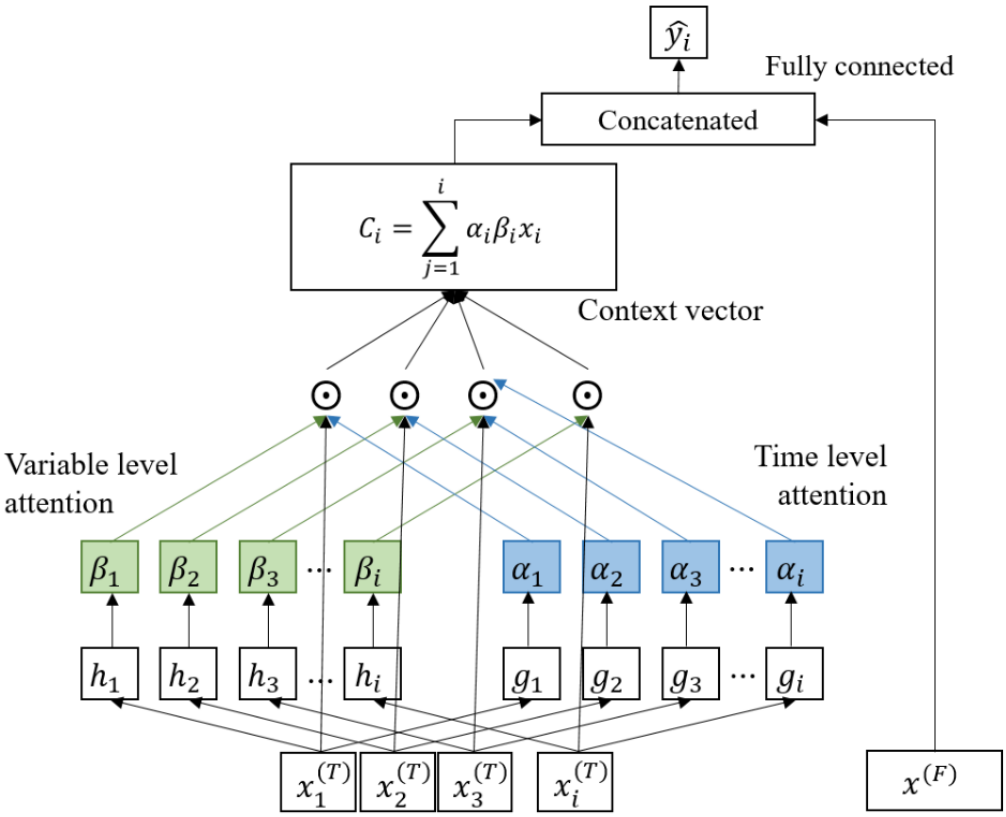


Development of a Weight Loss Prediction Model

We developed an interpretable RNN model by modifying the reverse time attention model (RETAIN) [16}. As the purpose of the RETAIN model is to predict the risk of cardiac disease using sequential data on categorical data such as diagnosis (International Statistical Classification of Diseases, Tenth Revision), we improved the model to (1) give leverage to continuous variables such as weight input frequency and time-fixed variables such as user profile and (2) convert this into a regression model for the prediction of weight after 16 weeks while preserving the linearity of the model for interpretability.

We modified the prediction model to (1) consider time-fixed variables by concatenating initial status as auxiliary inputs, (2) use continuous variables by removing the embedding layer, and (3) convert this into a regression model by not using the activation function for classification (Figure 2).

Figure 2. Model architecture.



Assuming we have N timeseries related to app usage with i length, where and k time-fixed variables where and a target series y of length i, where $\mathbf{y}=[y_1,\ldots,y_i]$, and . By stacking N timeseries, we define a multivariate input series and contemporary multivariate input vector is denoted at time i by . Given X_i and $x^{(F)}$, we leverage 2 types of data for our model to predict contemporary target values using deep learning , namely . The concatenation operation is denoted by $\oplus$ and element-wise multiplication is denoted by $\odot$. $A_{:i}$ is denoted by slice 1-dimensional tensor A from first index to ith index (Multimedia Appendix 1).

We attempted to predict the participant's weight after a 16-week intervention using the output vector $y_i \in \{0, \infty\}$ by concatenating c_i from time-variant data and $x^{(F)}$ and applying a linear transformation with w weight vector, where $w \in$. This is expressed as follows:

where c_i denotes the context vector $c_i \in$ and is the sum of the time-level variables weighted by attention α_i, β_i. Time-level attention weighting α_i is a scalar value that reflects the relative importance of the data at time i, and it ranges from 0 to 1. β_i is a vector with N length that explains the importance of each value of within i time.

Consistent with RETAIN, interpretation of the weight prediction model involves getting the time-level attention weighting and variable-level attention weighting from each RNN g_i, h_i [16]. Therefore, the equation can be rewritten, and the contributions of the predicted value of the model can be calculated as follows:

Therefore, the contribution of the Nth time-variant variable $\omega(\hat{y}_i, x_{jN})^{(T)}$ and that of the mth time-fixed variable $\omega(\hat{y}_i, x_m)^{(F)}$ can also be written as

where each contribution coefficient of the time-variant variable is , and each contribution coefficient of the time-fixed variable is .

We subsequently generated 3 dimensions of time-variant variables based on app usage patterns (window, week, variables) to provide immediate feedback if users performed behaviors that infringed on weight loss from week to week, and to improve model performance, and for real-time prediction of future target value, we fixed all elements of y as the weight at 16 weeks.

Furthermore, we leveraged information on the shape of the weight loss trend week to week. We used shape-based timeseries clustering on the weight logs; individuals' weights were assigned into clusters at each time for real-time prediction. The weight loss trajectory can be labeled at each time point and can represent the historical shape of weight loss, not actual weight (kilograms) from the first week to each subsequent week. This trajectory does not require the outcome of the model. Therefore, the real-time model can leverage a user's historical shape of weight loss at each time. To exclude the fact that weight (the outcome of the model) is directly used as a feature, we normalized each weight log by converting their means to 0, and the variance to 1. Then, we used the clustering number in the prediction process. Dynamic time warping was used to measure the distance between user weight loss trajectories; we used *k*-means clustering with this metric. We found 5 to be the optimal number of clusters. For individuals, a weight loss cluster was assigned for each time. An example of weight loss is included in Multimedia Appendix 2.

Model Evaluation With Cross-Validation Procedure

The performance of the regression model was evaluated using the mean absolute percentage error (MAPE) [23]. MAPE was calculated for the entire intervention (from week 1 to week 16) to identify model accuracy. Furthermore, to evaluate the consistency of the model's performance, 5-fold cross-validation (training set: 70% and validation set: 30%) was performed.

Identification of Contribution Coefficients

We identified the contribution coefficients by exploring the time-level attention weighting and variable-level attention weighting for a window of 16 weeks to ascertain the contributions of week 1 to week 16 in predicting weight. Furthermore, we explored the attention weighting of the predictive model by calculating the average of the attention weightings used for predicting weight in test data.

Ethics

The study was approved by the institutional review board at Advarra (CR00123125). The anonymous and deidentified nature of the retrospective log data made obtaining informed consent unnecessary.

Results

A total of 17,867 eligible users were included in this study. Their average age was 43.2 years (SD 10.8), and almost all users were female (16,460/17,867, 92.2%). The mean initial BMI was 33.7 kg/m^2, and users lost a mean of 4.6 kg (SD 4.8). Compared to their initial weight, a majority (5908/17,867 33.1%) of users lost 5% to 10% of their initial weight, followed by users who lost 2% to 5% of their initial weight (4894/17,867, 27.4%) (Table 1).

Table 1. Demographic characteristics of eligible users.

Variables	Eligible users (n=17,867)
Age (years), mean (SD)	43.2 (10.8)
Gender, n (%)	
Female	16,470 (92.2)
Male	1397 (7.8)
Height (cm), mean (SD)	166.4 (7.5)
Initial weight (kg), mean (SD)	93.4 (17.8)
Initial BMI (kg/m^2), mean (SD)	33.7 (5.8)
Last weight (kg), mean (SD)	88.8 (17.3)
Weight loss[a] (kg), mean (SD)	4.6 (4.8)
Outcome, n (%)	
Gained >2%	693 (3.9)
Stable	4075 (22.8)
Loss of 2%-5%	4894 (27.4)
Loss of 5%-10%	5908 (33.1)
Loss of 10%-15%	1983 (11.1)
Loss >15%	314 (1.8)

[a]Weight loss = initial weight – final weight.

The overall MAPE of the model was 3.50%. One, 8, 14, and 16 weeks after the coaching program, the model MAPEs were 3.78%, 3.45%, 3.41%, and 3.45%, respectively. Furthermore, the results of cross-validation showed that the error rates of the predicted value decreased as coaching weeks went by in folds 2, 4, and 5 (Figure 3). However, in folds 1 and 3, the MAPE increased to 3.45%.

The overall time-level attention weightings showed that the weight per time was almost evenly distributed as 0.0625; however, the attention weightings gradually decreased over time toward the last week of the coaching program (Figure 4A). The variable level attention weightings of each variable remained at 1 or –1 consistently for 16 weeks; however, the cluster assignment for yo-yo trajectories started off as negative during the first 8 weeks, then became positive in the subsequent 8 weeks (Figure 4B).

The time-variant weightings, including weight input frequency, meal input adherence, and exercise input frequency, had values of –0.340, –0.372, and 0.248, respectively. Among the cluster assignments, a sharp decrease in weight loss trajectory had a weighting of 1.066. Among the time-fixed variables, initial weight highly contributed to predicting weight (0.935). Furthermore, users whose initial weight was in the overweight range were expected to have a lower weight (–0.085). Those whose weights were from obese class I, class II, and class III were expected to have higher weights (0.084, 0.072, and 0.097, respectively) (Table 2).

For the contribution coefficient, weight input frequency, meal input adherence, steps, exercise input, and a sharply decreased weight loss trajectory had negative contribution coefficients. Among them, the sharply decreased trajectory had the most negative contribution coefficient (–0.066), followed by the weight input frequency, meal input frequency, and exercise input frequency (–0.021, –0.0232, and –0.0155, respectively). The caloric intake per kilocalorie had a positive contribution coefficient (0.098). Among the trajectories, assignment into the yo-yo group had a negative contribution coefficient (–0.0059) in the first week, however, this became a positive coefficient after 8 weeks (Figure 5).

Figure 3. Mean absolute percentage error (MAPE) of model in each 5-fold cross-validation.

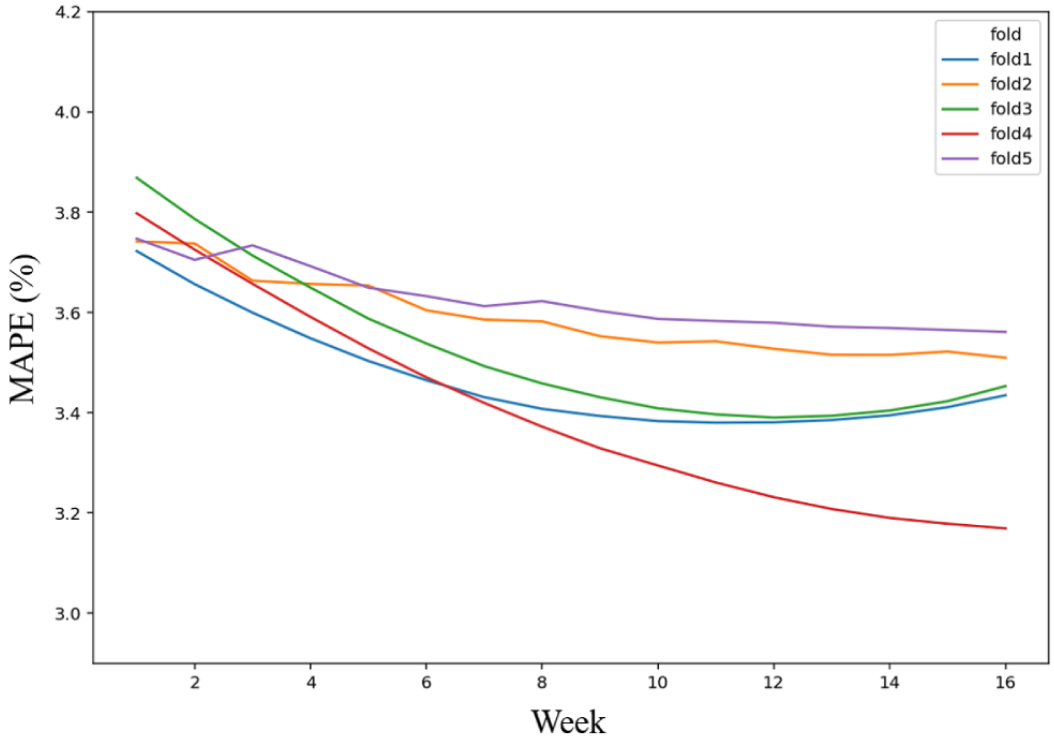


Figure 4. (A) Time-level attention α and (B) variable-level attention β of each variable for the 16-week coaching program.

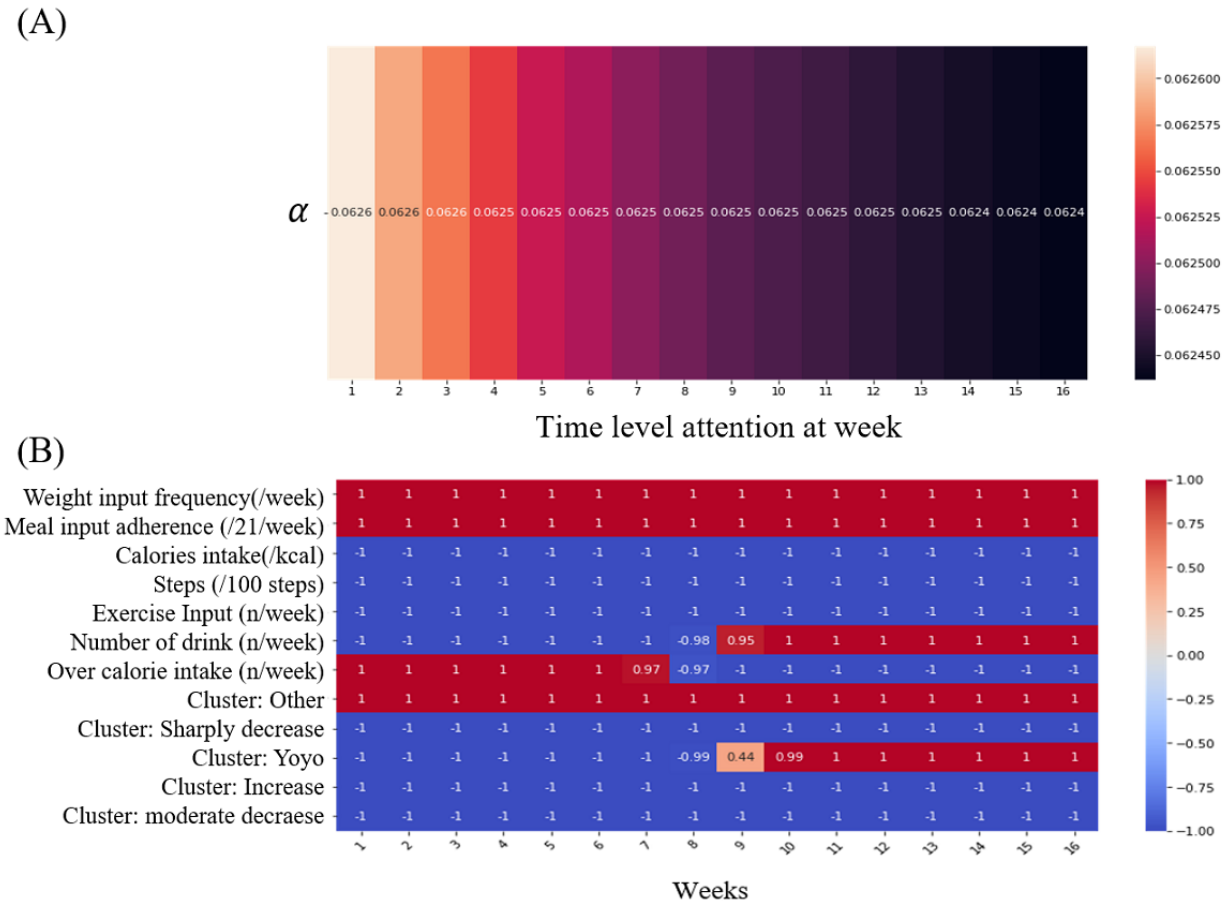
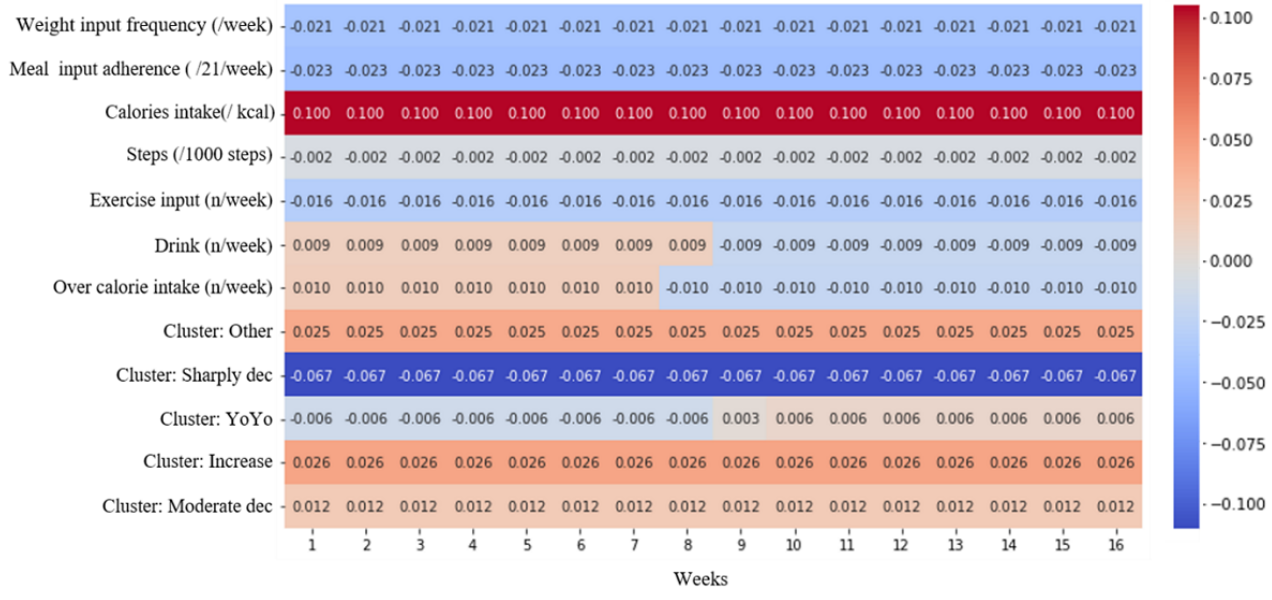


Table 2. Weightings for time-variant and time-fixed variables in matrix w.

Variables	Weightings
Time-variant	
Weight input frequency (n/week)	–0.340
Meal input adherence (%/week)	–0.372
Exercise inputs frequency (n/week)	0.248
Calorie intake (kcal/per days)	–1.597
Steps (/1000 daily steps)	0.031
Alcohol drink (n/week)	–0.150
Over-calorie intake (n/week)	0.164
Weight loss trajectories (if assigned)	
Sharp decrease	1.066
Moderate decrease	–0.188
Yo-yo	0.094
Increase	–0.422
Other	0.395
Time-fixed	
Gender	
Female	0.160
Male	–0.091
Age (years)	0.002
Height (cm)	0.013
Initial weight (kg)	0.935
Obesity class (if assigned)	
Overweight	–0.085
Obese class I	0.084
Obese class II	0.072
Obese class III	0.097

Figure 5. Contribution coefficients for predicting weight change in each week for each variable.



Discussion

Principal Results

In this study, a transparent RNNs model was constructed utilizing both timeseries and time-fixed data for weight loss prediction after a mobile-based intervention targeting obesity. Currently, the explanation methods based on attention mechanisms are referred to as explaining individual predictions because attention can be generated in each person by providing a contributing role in prediction. However, our research first demonstrated the connection between individuals and a global pattern with homogenous length timeseries. Through the identification of time-level and variable-level attention weighting, we identified that weight, meal, and exercise inputs are important factors for weight loss in the context of mobile apps. Furthermore, the findings regarding time-level attention weighting may suggest that constant use of the app is important in mobile-based interventions. In individual profiles, we identified that being male and being overweight were relatively negative contributing factors to weight after the mobile program intervention.

In addition, this predictive model considered both time-variant variables, such as weekly adherence to the app, and time-fixed variables, such as the initial status of users to predict week-to-week weight change. Recently, interpretable RNN architecture has been proposed in machine learning areas [16,24-26]. However, the simple application of this architecture may be difficult in research problems involving classification or regression, such as those in medical fields where demographic information is important for prediction. To address this limitation, we modified the RETAIN model while preserving model transparency using conditional RNNs by adding auxiliary inputs such as basic patient information. Our results, which predicted weight based on interpretation of both demographic status and app usage patterns, showed how each variable was important in predicting weight change each week.

Finally, based on the lifelog data of approximately 17,000 users, we found that a deep learning model that takes advantage of the sequential nature of lifelog data was able to perform well in the general population. Research on models for weight loss prediction has been conducted in fields such as bariatrics or pharmacology on small populations; therefore, it is difficult to generalize the findings of these previous studies [12,27] to the general population. Since we used a general population for this study, these findings may be applied to a general and heterogeneous group.

Interpretable Algorithm

It is important for end-users such as the patient or clinician to interpret and understand the AI system because erroneous predictions are especially risky and expensive [28]. As black-box machine learning models are frequently employed due to the importance of prediction in critical contexts (such as within the medical field), the demand for transparency in AI, from end-users to research staff, is increasing [29,30]. This has led to a strong interest in explainable AI, which can provide details or reasons to make its functioning clear or easy to understand [17]. Explainable AI research to make RNNs transparent has been conducted [16,24-26]; however, it is difficult to directly adopt these algorithms for problems in medical fields without patient profiles because time-fixed variables have not been properly handled. Therefore, we modified the interpretable AI model to handle both types of data while preserving transparency, and our work can be a cornerstone of medical research in discovering new insights. In addition, contribution coefficients from the model can provide intuitive insight into why the model's predicted values differed compared to actual data of users' behaviors and weight change from week to week [16].

Timely Feedback and Contributing Factors to Weight Loss

We found that weighting was equally distributed each week with a value of 0.0625 (Figure 4); however, this indicated that the initial usage of the app was slightly more important than later usage (from 0.0626 to 0.0624). Given that time-level attention weighting is relative, and the sum of these is 1, the importance of each week is almost equivalent because the weight loss program continued for 16 weeks ($0.0625 \fallingdotseq 1/16$). Therefore, this implies that each week in the diet program is nearly equally important, but early usage is relatively more important. In addition, timely feedback may be important if a user behaves in a way that is detrimental to weight loss for a week. To address this, our model was developed to cover 3 types of sequence data (window, week, variables) to forecast changes from any week in the mobile program. Therefore, our model can provide timely feedback by predicting weight change from week to week.

Most of the variable-level attention weightings shown were either 1 or –1 (Figure 5). Each attention weighting needs to be understood to infer why the variable-level attention is represented as –1 or 1. The time-level attention weighting was always positive as it is derived from the softmax function, and the variable-level attention weighting ranged from –1 to 1 as it is derived from the hyperbolic tangent function, which is multiplied with the parameter w in linear transformation; therefore, we inferred that variable-level attention represented the sign of weighting at each time step in the model. Instead, parameter w in linear transformation seemed to represent the magnitude of weightings, and its values were constant regardless of the time step. Therefore, to at least understand both the magnitude and sign of the weightings (eg, regardless of the time step), variable-level attention weighting and W should be multiplied together.

Weight and meal inputs core behavior components of self-monitoring showed results consistent with those of previous studies [31-33]. A previous study [34] that did not consider timeseries identified factors contributing to weight loss using multivariate and univariate regression consistent with our findings; overall weight, meal input, and exercise input had a significant correlation with weight loss [34]. Similarly, calories were also a contributing factor to weight gain [34]. In our study, the weight, meal, and exercise inputs had consistent contribution coefficients of –0.021, –0.023, and –0.016, respectively, while caloric intake had a contribution coefficient of 0.100; these were found using interpretable RNNs, though our model cannot provide the statistical significance of contribution coefficient.

Furthermore, the yo-yo weight loss trajectory had a negative contribution coefficient of –0.006 for 8 weeks, then changed to a positive value of 0.06 at 8 weeks. To determine whether the weight loss trajectory is classified as a yo-yo or decreased pattern, the eighth week may be a critical week. In other words, to avoid a yo-yo trajectory, self-monitoring using the mobile app may be important especially for the initial 8 weeks. Meanwhile, because this model uses the historical shape of weight loss, there may be a concern that the model is weak because it indirectly learns the weight trajectory shape. Thus, we also conducted an experiment where the model without weight loss trajectories was retrained, and the result showed that the model maintained the robustness of performance (Multimedia Appendix 3). From the user's weight entry, we derived and leveraged 2 variables—weight loss trajectory and weight input. In the real world, although coaches encourage users to enter their weight, there may be users without weight records. For users without any weight records, the accuracy and interpretation of the model's prediction are limited. From the comparison between the model with and without weight loss trajectory as the feature (Multimedia Appendix 3), the trajectories may not be a high-contributing factor for predicting weight loss because the model compensates. Despite this, adding weight loss trajectory may provide a hint for early intervention with cluster assignment. For example, it is possible to coach the user by considering the user's weight loss trajectory compared to the weight loss plateau or rebounding shape. Therefore, the selection of feature needs to be considered with the interpretability of each feature in mind.

Unexpectedly, drinking frequency and overall caloric intake did not make a consistent contribution to the model. This may be due to low retention, causing the number of drinks and overconsumption of calories not to be recorded because the records related to meal logs were not entered consistently within the 16 weeks.

When AI is used as a decision aid (eg, for a coach), an explanation can improve its trustworthiness [35]. For example, a coach (or another health provider) may highlight user's behaviors that cause weight gain based on the interpretation of the model. Based on the feedback of intervention theory, feedback focused on the behavior task itself introduces the importance of timing in feedback delivery [36,37]. Also, a feedback strategy based on such a theory should be personalized and goal oriented. From this point of view, users may modify their behavior to increase weight loss based on the results of a real-time prediction and contribution factors, which improves their motivation [36,37].

Limitations

Our study has some inherent limitations due to its implementation of interpretable AI using lifelog data. The individuals included in this study had purchased a subscription to a weight loss program and were selected if they had weight records for 16 weeks. This may result in selection bias wherein the chosen individuals were those with high adherence or the intention to lose weight. However, given that most of the participants who used other weight loss apps were in their forties, and a high rate of users were female in previous studies, characteristics similar to those in our study, our study is able to show robust findings with respect to population group that uses these apps [20,38-42]. Furthermore, given that the average 30-day retention rate is 3.4% for health and fitness apps in the United States, and that some researchers have reported a 3% to 8% 30-day retention rate for mental health apps, this app has an inherently higher retention rate (completion of 16-week program) compared to peer health care apps [43-45].

Despite this, there is limited research on predictive models for weight change using lifelog data and identification of contributing factors to weight loss. These results can make up for the lack of evidence on the prediction of weight change in an electronically delivered intervention setting and contribute to better outcomes. Furthermore, the implementation of the model was based on a single mobile app for obesity management. Therefore, for more generalized evidence, research using lifelogs from various apps is needed.

Conclusion

In conclusion, we have demonstrated that an interpretable RNN can consider app usage behavior and the users' demographic characteristics while maintaining model transparency.

Acknowledgments

This study was supported by the Foundational Technology Development Program (NRF2019M3E5D406468221) of the Ministry of Science and Information Communication Technology, Republic of Korea

Conflicts of Interest

None declared.

Multimedia Appendix 1
Illustrated explanation of conditional RNN and explainability.
[DOCX File , 230 KB - mhealth_v9i3e22183_app1.docx]

Multimedia Appendix 2
Clustered weight loss trajectories using k-means with dynamic time warping.
[DOCX File , 255 KB - mhealth_v9i3e22183_app2.docx]

Multimedia Appendix 3
Model performance from 5 cross-validations after retraining the model without trajectory features.
[DOCX File , 49 KB - mhealth_v9i3e22183_app3.docx]

References

1. Roberto CA, Swinburn B, Hawkes C, Huang TT, Costa SA, Ashe M, et al. Patchy progress on obesity prevention: emerging examples, entrenched barriers, and new thinking. Lancet 2015 Jun 13;385(9985):2400-2409. [doi: 10.1016/S0140-6736(14)61744-X] [Medline: 25703111]
2. Ford ND, Patel SA, Narayan KMV. Obesity in low- and middle-income countries: burden, drivers, and emerging challenges. Annu Rev Public Health 2017 Mar 20;38:145-164. [doi: 10.1146/annurev-publhealth-031816-044604] [Medline: 28068485]
3. Apovian CM. The obesity epidemic — understanding the disease and the treatment. N Engl J Med 2016 Jan 14;374(2):177-179. [doi: 10.1056/nejme1514957]
4. Fruh SM. Obesity: risk factors, complications, and strategies for sustainable long-term weight management. J Am Assoc Nurse Pract 2017 Oct;29(S1):S3-S14 [FREE Full text] [doi: 10.1002/2327-6924.12510] [Medline: 29024553]
5. Yusefzadeh H, Rashidi A, Rahimi B. Economic burden of obesity: a systematic review. Soc Health Behav 2019;2(1):7. [doi: 10.4103/shb.shb_37_18]
6. López G, González I, Jimenez-Garcia E, Fontecha J, Brenes JA, Guerrero LA, et al. Smart device-based notifications to promote healthy behavior related to childhood obesity and overweight. Sensors (Basel) 2018 Jan 18;18(1) [FREE Full text] [doi: 10.3390/s18010271] [Medline: 29346302]
7. Alnuaimi A, Rawaf S, Hassounah S, Chehab M. Use of mobile applications in the management of overweight and obesity in primary and secondary care. JRSM Open 2019 Apr 05;10(3):205427041984382. [doi: 10.1177/2054270419843826]
8. Nikolaou CK, Lean MEJ. Mobile applications for obesity and weight management: current market characteristics. Int J Obes (Lond) 2016 Nov 22. [doi: 10.1038/ijo.2016.186] [Medline: 27780974]
9. Deering M. Issue brief: Patient-generated health data and health IT. The Office of the National Coordinator for Health Information Technology. 2013. URL: https://www.healthit.gov/sites/default/files/pghd_brief_final122013.pdf [accessed 2021-02-26]
10. Kim H, Lee S, Kim JH. Real-world evidence versus randomized controlled trial: clinical research based on electronic medical records. J Korean Med Sci 2018 Aug 20;33(34):e213 [FREE Full text] [doi: 10.3346/jkms.2018.33.e213] [Medline: 30127705]
11. Gollamudi SS, Topol EJ, Wineinger NE. A framework for smartphone-enabled, patient-generated health data analysis. PeerJ 2016;4:e2284 [FREE Full text] [doi: 10.7717/peerj.2284] [Medline: 27547580]
12. Wise ES, Hocking KM, Kavic SM. Prediction of excess weight loss after laparoscopic Roux-en-Y gastric bypass: data from an artificial neural network. Surg Endosc 2015 May 28;30(2):480-488. [doi: 10.1007/s00464-015-4225-7]
13. Dugan TM, Mukhopadhyay S, Carroll A, Downs S. Machine learning techniques for prediction of early childhood obesity. Appl Clin Inform 2017 Dec 19;06(03):506-520. [doi: 10.4338/aci-2015-03-ra-0036]
14. Hammond R, Athanasiadou R, Curado S, Aphinyanaphongs Y, Abrams C, Messito MJ, et al. Predicting childhood obesity using electronic health records and publicly available data. PLoS ONE 2019 Apr 22;14(4):e0215571. [doi: 10.1371/journal.pone.0215571]
15. Kim J, Park YR, Lee JH, Lee J, Kim Y, Huh JW. Development of a real-time risk prediction model for in-hospital cardiac arrest in critically ill patients using deep learning: retrospective study. JMIR Med Inform 2020 Mar 18;8(3):e16349. [doi: 10.2196/16349]
16. Choi E, Bahadori M, Sun J, Kulas J, Schuetz A, Stewart W. RETAIN: an interpretable predictive model for healthcare using reverse time attention mechanism. 2016 Presented at: Advances in Neural Information Processing Systems 29 (NIPS 2016); December 9-10; Barcelona, Spain.
17. Samek W, Montavon G, Vedaldi A, Hansen LK, Müller KR, editors. Explainable AI: Interpreting, Explaining and Visualizing Deep Learning (Vol. 11700). Cham: Springer Nature; 2019.
18. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 2020 Jun;58:82-115. [doi: 10.1016/j.inffus.2019.12.012]
19. Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. Wiley Interdiscip Rev Data Min Knowl Discov 2019;9(4):e1312 [FREE Full text] [doi: 10.1002/widm.1312] [Medline: 32089788]
20. Kim Y, Oh B, Shin H. Effect of mHealth with offline antiobesity treatment in a community-based weight management program: cross-sectional study. JMIR Mhealth Uhealth 2020 Jan 21;8(1):e13273 [FREE Full text] [doi: 10.2196/13273] [Medline: 31961335]
21. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 2015 Jan 07;350(jan07 4):g7594-g7594. [doi: 10.1136/bmj.g7594]

22. Defining adult overweight and obesity. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/obesity/adult/defining.html [accessed 2021-02-21]
23. de Myttenaere A, Golden B, Le Grand B, Rossi F. Mean absolute percentage error for regression models. Neurocomputing 2016 Jun;192:38-48. [doi: 10.1016/j.neucom.2015.12.114]
24. Foerster J, Gilmer J, Sohl-Dickstein J, Chorowski J, Sussillo D. nput switched affine networks: an RNN architecture designed for interpretability. In: Proceedings of the 34th International Conference on Machine Learning Volume 70. 2017 Presented at: 34th International Conference on Machine Learning; Aug 11-15; Sydney, Australia.
25. Guo T, Lin T. Exploring interpretable lstm neural networks over multivariable data. 2019 Presented at: International Conference on Machine Learning; June 9-15; Long Beach, California, United States.
26. Qin Y, Song D, Chen H, Cheng W, Jiang G. A dual-stage attention-based recurrent neural network for time series prediction. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. 2017 Presented at: International Joint Conference on Artificial Intelligence; August 19-25; Melbourne, Australia p. 2627-2633. [doi: 10.24963/ijcai.2017/366]
27. Langius JAE, Twisk J, Kampman M, Doornaert P, Kramer MHH, Weijs PJM, et al. Prediction model to predict critical weight loss in patients with head and neck cancer during (chemo)radiotherapy. Oral Oncol 2016 Jan;52:91-96. [doi: 10.1016/j.oraloncology.2015.10.021] [Medline: 26564309]
28. Ahmad M, Eckert C, Teredesai A. Interpretable machine learning in health care. In: Proceedings of the ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics. 2018 Presented at: International Conference on Bioinformatics, Computational Biology, and Health Informatics; August 29 - September 1; Washington DC p. 559-560. [doi: 10.1145/3233547.3233667]
29. Preece A, Harborne D, Braines D, Tomsett R, Chakraborty S. Stakeholders in explainable AI. 2018 Presented at: AAAI FSS-18: Artificial Intelligence in Government and Public Sector; October 18–20; Arlington, Virginia, USA.
30. Ge W, Huh J, Park Y, Lee J, Kim Y, Turchin A. An interpretable ICU mortality prediction model based on logistic regression and recurrent neural networks with LSTM units. 2018 Presented at: AMIA Annual Symposium Proceedings; November 3-7; San Francisco, California, United States.
31. Gokee-Larose J, Gorin AA, Wing RR. Behavioral self-regulation for weight loss in young adults: a randomized controlled trial. Int J Behav Nutr Phys Act 2009;6:10 [FREE Full text] [doi: 10.1186/1479-5868-6-10] [Medline: 19220909]
32. Patel ML, Hopkins CM, Brooks TL, Bennett GG. Comparing self-monitoring strategies for weight loss in a smartphone app: randomized controlled trial. JMIR Mhealth Uhealth 2019 Feb 28;7(2):e12209 [FREE Full text] [doi: 10.2196/12209] [Medline: 30816851]
33. Khaylis A, Yiaslas T, Bergstrom J, Gore-Felton C. A review of efficacious technology-based weight-loss interventions: five key components. Telemed J E Health 2010 Nov;16(9):931-938 [FREE Full text] [doi: 10.1089/tmj.2010.0065] [Medline: 21091286]
34. Chin SO, Keum C, Woo J, Park J, Choi HJ, Woo JT, et al. Successful weight reduction and maintenance by using a smartphone application in those with overweight and obesity. Sci Rep 2016 Nov 07;6:34563 [FREE Full text] [doi: 10.1038/srep34563] [Medline: 27819345]
35. Wang D, Yang Q, Abdul A, Lim B. Designing theory-driven user-centric explainable AI. 2019 Presented at: CHI conference on human factors in computing systems; May 4-9; Glasgow, Scotland. [doi: 10.1145/3290605.3300831]
36. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2016 Sep 23. [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
37. Schembre SM, Liao Y, Robertson MC, Dunton GF, Kerr J, Haffey ME, et al. Just-in-time feedback in diet and physical activity interventions: systematic review and practical design framework. J Med Internet Res 2018 Mar 22;20(3):e106 [FREE Full text] [doi: 10.2196/jmir.8701] [Medline: 29567638]
38. Krukowski RA, Harvey-Berino J, Ashikaga T, Thomas CS, Micco N. Internet-based weight control: the relationship between web features and weight loss. Telemed J E Health 2008 Oct;14(8):775-782 [FREE Full text] [doi: 10.1089/tmj.2007.0132] [Medline: 18954247]
39. Alexander E, Tseng E, Durkin N, Jerome GJ, Dalcin A, Appel LJ, et al. Factors associated with early dropout in an employer-based commercial weight-loss program. Obes Sci Pract 2018 Dec;4(6):545-553 [FREE Full text] [doi: 10.1002/osp4.304] [Medline: 30574348]
40. Serrano KJ, Yu M, Coa KI, Collins LM, Atienza AA. Mining health app data to find more and less successful weight loss subgroups. J Med Internet Res 2016 Jun 14;18(6):e154 [FREE Full text] [doi: 10.2196/jmir.5473] [Medline: 27301853]
41. Painter SL, Ahmed R, Kushner RF, Hill JO, Lindquist R, Brunning S, et al. Expert coaching in weight loss: retrospective analysis. J Med Internet Res 2018 Mar 13;20(3):e92 [FREE Full text] [doi: 10.2196/jmir.9738] [Medline: 29535082]
42. He C, Wu S, Zhao Y, Li Z, Zhang Y, Le J, et al. Social media-promoted weight loss among an occupational population: cohort study using a WeChat mobile phone app-based campaign. J Med Internet Res 2017 Oct 23;19(10):e357 [FREE Full text] [doi: 10.2196/jmir.7861] [Medline: 29061555]
43. Farago P. App engagement: the matrix reloaded. Flurry. 2012. URL: https://www.flurry.com/post/113379517625/app-engagement-the-matrix-reloaded [accessed 2021-03-10]

44. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
45. Tison G, Hsu K, Hsieh J, Ballinger B, Pletcher M, Marcus GM, et al. Achieving high retention in mobile health research using design principles adopted from widely popular consumer mobile apps. Circulation 2017;136(suppl_1):A21029 [FREE Full text]

Abbreviations

AI: artificial intelligence
MAPE: mean absolute percentage error
RNN: recurrent neural network
RETAIN: reverse time attention model

Edited by G Eysenbach, R Kukafka; submitted 17.08.20; peer-reviewed by B Spring; comments to author 09.09.20; revised version received 15.10.20; accepted 17.01.21; published 29.03.21.

Please cite as:
Kim HH, Kim Y, Park YR
Interpretable Conditional Recurrent Neural Network for Weight Change Prediction: Algorithm Development and Validation Study
JMIR Mhealth Uhealth 2021;9(3):e22183
URL: https://mhealth.jmir.org/2021/3/e22183
doi:10.2196/22183
PMID:33779574

Original Paper

A Screening Method Using Anomaly Detection on a Smartphone for Patients With Carpal Tunnel Syndrome: Diagnostic Case-Control Study

Takafumi Koyama[1], MD; Shusuke Sato[2], BEng; Madoka Toriumi[2], MEng; Takuro Watanabe[2], MEng; Akimoto Nimura[3], MD, PhD; Atsushi Okawa[1], MD, PhD; Yuta Sugiura[2], PhD; Koji Fujita[3], MD, PhD

[1]Department of Orthopedic and Spinal Surgery, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo, Japan
[2]School of Science for Open and Environmental Systems, Graduate School of Science and Technology, Keio University, Kanagawa, Japan
[3]Department of Functional Joint Anatomy, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo, Japan

Corresponding Author:
Koji Fujita, MD, PhD
Department of Functional Joint Anatomy
Graduate School of Medical and Dental Sciences
Tokyo Medical and Dental University
1-5-45, Yushima, Bunkyo-ku
Tokyo, 1138519
Japan
Phone: 81 3 5803 5279
Email: fujiorth@tmd.ac.jp

Abstract

Background: Carpal tunnel syndrome (CTS) is a medical condition caused by compression of the median nerve in the carpal tunnel due to aging or overuse of the hand. The symptoms include numbness of the fingers and atrophy of the thenar muscle. Thenar atrophy recovers slowly postoperatively; therefore, early diagnosis and surgery are important. While physical examinations and nerve conduction studies are used to diagnose CTS, problems with the diagnostic ability and equipment, respectively, exist. Despite research on a CTS-screening app that uses a tablet and machine learning, problems with the usage rate of tablets and data collection for machine learning remain.

Objective: To make data collection for machine learning easier and more available, we developed a screening app for CTS using a smartphone and an anomaly detection algorithm, aiming to examine our system as a useful screening tool for CTS.

Methods: In total, 36 participants were recruited, comprising 36 hands with CTS and 27 hands without CTS. Participants controlled the character in our app using their thumbs. We recorded the position of the thumbs and time; generated screening models that classified CTS and non-CTS using anomaly detection and an autoencoder; and calculated the sensitivity, specificity, and area under the curve (AUC).

Results: Participants with and without CTS were classified with 94% sensitivity, 67% specificity, and an AUC of 0.86. When dividing the data by direction, the model with data in the same direction as the thumb opposition had the highest AUC of 0.99, 92% sensitivity, and 100% specificity.

Conclusions: Our app could reveal the difficulty of thumb opposition for patients with CTS and screen for CTS with high sensitivity and specificity. The app is highly accessible because of the use of smartphones and can be easily enhanced by anomaly detection.

(JMIR Mhealth Uhealth 2021;9(3):e26320) doi:10.2196/26320

KEYWORDS

carpal tunnel syndrome; anomaly detection; machine learning; smartphone; screening; thumb; diagnostic; data collection; app; algorithm

Introduction

Carpal tunnel syndrome (CTS) is a medical condition caused by compression of the median nerve in the carpal tunnel due to aging or hand overuse [1]. Patients with CTS develop numbness from the thumb to the ring finger and, in severe cases, thenar muscle atrophy [2]. Because the symptoms impair thumb motions, CTS can impede everyday movements, such as holding a pen or chopsticks and handling buttons on clothes [3,4]. The prevalence of CTS is approximately 2% to 14%, and it affects more women than men [5,6]. Since most patients with CTS are aged 40 years or older [5,6] and the number of older people is increasing worldwide, the number of patients with CTS is expected to increase. Nonsurgical therapy, such as a wrist brace [7] or steroid injection into the carpal tunnel [8], is typically prescribed, but surgery is often necessary for severe symptoms [2]. Patients often delay seeking medical attention until the numbness worsens and thenar atrophy develops. The symptoms in severe cases recover slowly postoperatively [9,10]; therefore, early diagnosis of CTS and surgery before the symptoms worsen is important.

Physical findings, such as the Tinel sign or Phalen test, may be used; however, their sensitivity and specificity are not high [11,12]. Although a nerve conduction study (NCS) is considered useful for diagnosing CTS [13,14], the equipment is expensive and the process can be painful and long (up to an hour). In addition, a skilled technician must perform the detailed NCS [15]. Due to impaired access to NCSs, diagnosis is largely performed subjectively by doctors in clinics and small hospitals in which there are neither hand surgeons nor specialized equipment, contributing to delayed diagnosis.

In recent years, cameras and sensors in mobile devices have become smaller and more sophisticated and can now measure the state of the user. Various studies have been conducted on the use of mobile devices to acquire physical information and diagnose diseases [16-18]. Fujita et al [19] developed an app for screening CTS with an accuracy of 83% using a tablet and support vector machine, utilizing a machine learning technique. However, as patients with CTS are mainly aged 40 years and older and the rate of use of tablets in this age group is not high [20,21], the system may be difficult to introduce. Furthermore, the machine learning for the binary classification used in the aforementioned app needs 2 data sets, both from healthy controls and from patients. Large data sets are needed to enhance machine learning; however, these are difficult to collect due to the low prevalence of CTS.

To address these concerns, we developed a screening app for CTS using a smartphone and an anomaly detection algorithm [22] because the usage rate of smartphones is higher than that of tablets [20,21] and anomaly detection algorithms need only easily collected data sets of healthy controls. We aimed to examine whether our system was a useful screening tool for CTS.

Methods

Recruitment

This study was approved by the Institutional Review Board of Tokyo Medical and Dental University. Written informed consent was provided by all participants.

We recruited 21 preoperative patients (36 hands) with CTS at the Tokyo Medical and Dental University Hospital as the CTS group and 15 healthy volunteers (27 hands) at an osteopathic clinic as the non-CTS group from July 2018 to May 2019. Experienced hand surgeons diagnosed CTS based on symptoms, physical findings such as the Tinel sign and Phalen test, x-ray images of the hands, and NCSs measured by Neuropack X1 (Nihon Kohden). Patients were classified based on the Bland classification [23]. Patients with a history of other hand injury or surgery, recurrence after release surgery of the carpal tunnel, positive imaging findings indicative of first carpometacarpal or thumb metacarpophalangeal osteoarthritis (which could affect thumb motion), or suspicion of a disease of the cervical spine were excluded. In the non-CTS group, volunteers were excluded if they had a history of wrist, hand, or finger disease, injury, or surgery; finger numbness; thumb pain; or positive physical findings of CTS.

App Design

We used a Huawei P10 Lite (Huawei Technologies) phone and developed the app using Unity software (Unity Technologies). We also created a finger guide, which was attached to the back of the smartphone to fix the position of the fingers other than the thumb (Figure 1). The guide consisted of a component created with a 3D printer and 3 binding bands. The length of the binding band could be adjusted to adapt to the participant's finger thickness.

Figure 1. A finger guide attached to the back of the smartphone to fix the position of the fingers other than the thumb.

In this app, the player controlled a rabbit character with their thumb and collected vegetables (carrots, radishes, or eggplants) that appeared on the screen (screen A in Figure 2). When vegetables were hidden by the thumb, broad markers that indicated the direction of the vegetable were also displayed. The vegetables appeared sequentially in 12 directions along a circle with a 2-cm radius in a random order (screens B and C in Figure 2). The vegetables appeared alternatively in one of 12 directions and at the center, and the user collected them in each direction and in the center in turn (Figure 3). If a vegetable was not collected within 5 seconds, it disappeared and then reappeared at another place. In the practice phase, vegetables appeared randomly in 4 directions. Subsequently, in the measurement phase, the game ended after 2 sequences of vegetable appearances in 12 directions. The participants played the game in the app twice. The position of the thumb and the time were recorded, and the average time, average velocity, and maximum velocity of the thumb movement for the 12 directions were calculated.

Figure 2. The images of the app. A rabbit character and vegetables were displayed in the green circle. Vegetables were located at the center or edge of the circle, and markers were also displayed when the vegetables were located at the edge (A). Vegetables appeared in 12 numbered directions, and the numbers were reversed depending on whether the player used the left (B) or right (C) hand.

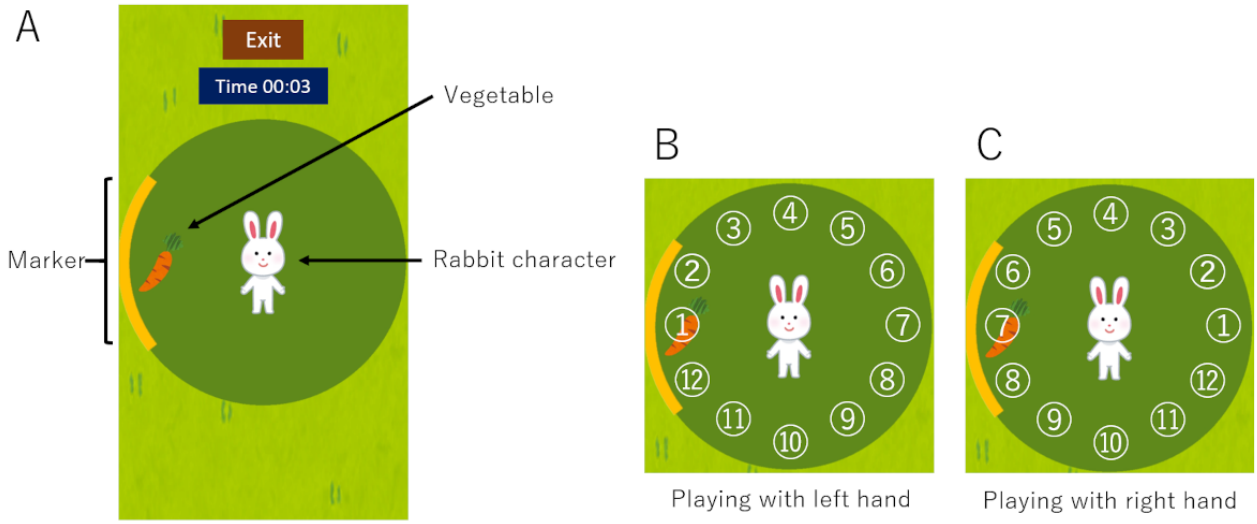
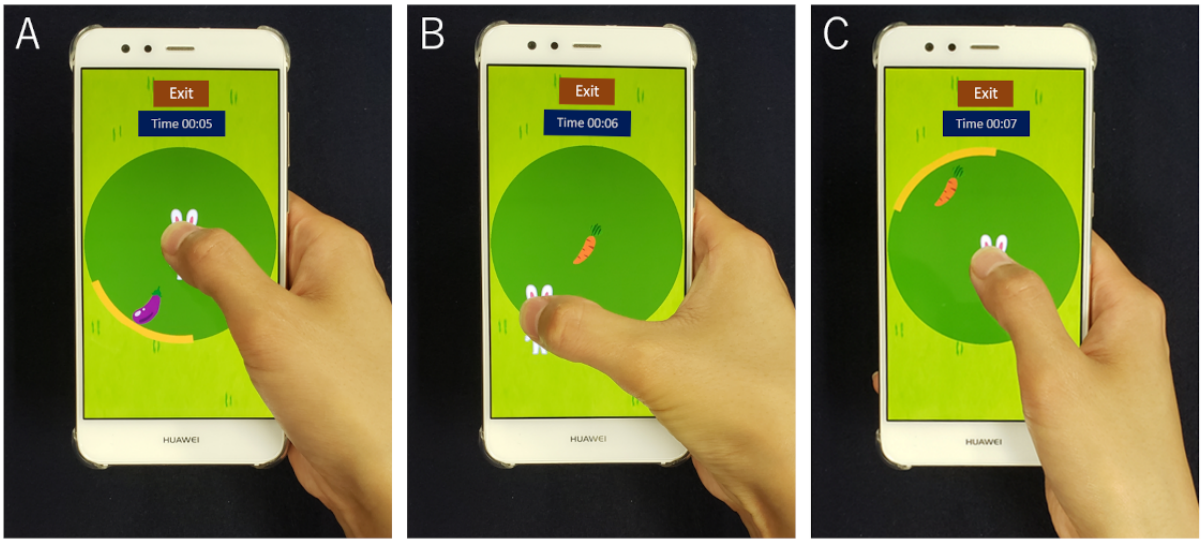


Figure 3. The images of the app while playing the game. The player touched and controlled a rabbit character with the thumb of each hand to collect vegetables. Vegetables appeared in one of 12 directions (A). When each vegetable was collected, the next appeared alternately at the center of the circle (B) or in another direction (C).

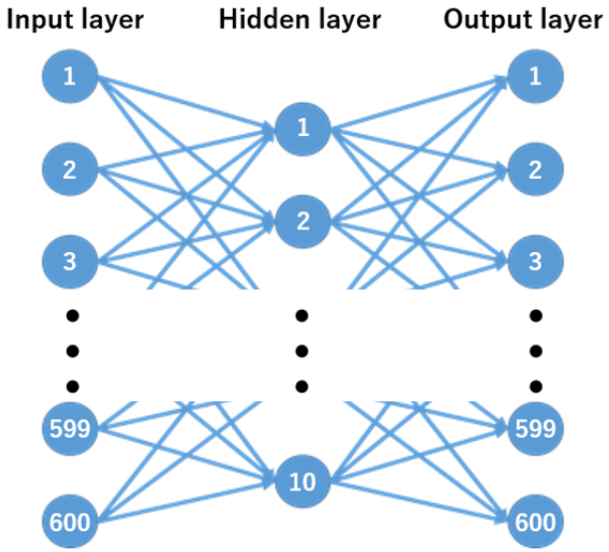


Statistical Analysis

We used 2-tailed Student *t* tests to compare the age of participants, average time, average velocity, and maximum velocity of the thumb movement for the 12 directions between the non-CTS and CTS groups. Chi-square tests were used to compare sex, playing side of the hand, and hand dominance between the non-CTS and CTS groups. A *P* value below .05 was considered to indicate statistical significance.

To generate a screening model that classified participants as CTS and non-CTS, we analyzed data sets using anomaly detection and an autoencoder (AE) [24,25]. Anomaly detection is the process of identifying data that differ from the norm in a data set. It has the advantage that it can be learned from normal group data only. The AE is a type of neural network with a 3-layer structure consisting of input, hidden, and output layers (Figure 4). The transformation of the input layer to the hidden layer is the encoder, and the attempted reconstruction of the hidden layer to the output layer constitutes the decoder. The AE performs unsupervised learning and is trained to reconstruct the input patterns. By reducing the number of units in a hidden layer compared with the number of units in the input layer, it enables dimensional compression.

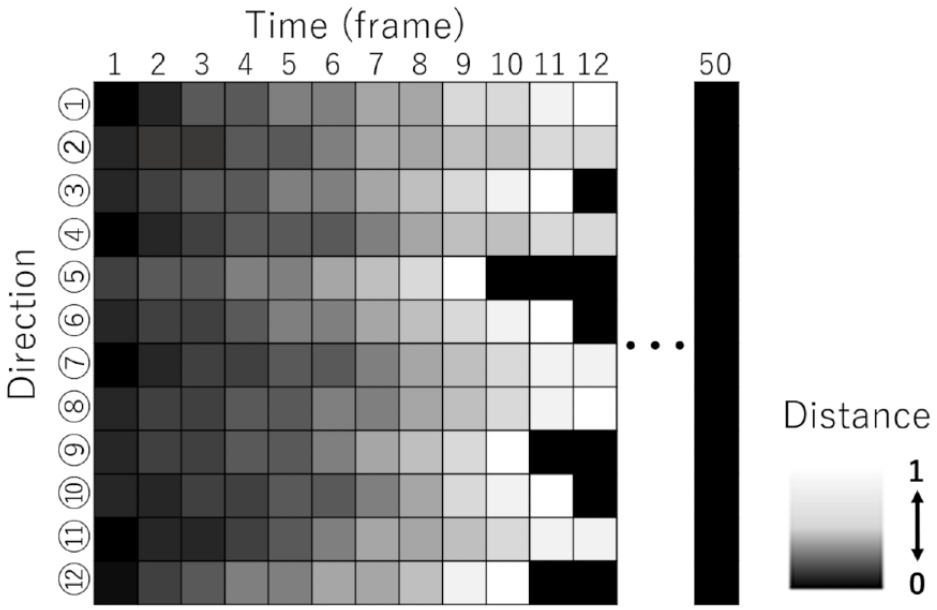
Figure 4. An image demonstrating how the autoencoder works. In our model, the input layer was 600 dimensions, the intermediate layer was 10 dimensions, and the output layer was 600 dimensions.



First, we calculated the distance to the center of the screen from the coordinate data and converted this into a value from 0 to 1. In our proposed model, the first lap was only played as practice for the participants to get used to the app, and only the second-lap data were used for the analysis. Next, a grayscale image was generated by arranging the pixel values with the vertical axis set as each direction and the horizontal axis set as time (Figure 5). The horizontal axis consisted of 5 seconds,

which is equal to 50 frames because the sampling rate was 10 Hz. Hence, the pixel count of the grayscale image was 600 pixels (12 directions × 50 frames). Finally, we validated the classification of non-CTS and CTS using the AE. For the AE, the 600 pixels of the grayscale image were used as the input layer, the intermediate layer was fixed at 10 dimensions, and the output layer was set to 600 dimensions (Figure 4).

Figure 5. Grayscale image generated by the pixel values with the vertical axis set as each direction and the horizontal axis set as time. The intensity of the pixel was defined by the distance between the thumb and the center; the greater the distance, the lighter the intensity. Pixels of the frames when the thumb reached the circumference (vegetables) were white, and all pixels to the right of the frames were set to be filled with black. The vertical axis was set as 12 directions and the horizontal axis was set at a fixed time (50 frames).



We used the data from 12 hands in the non-CTS group for the training of the AE and validated them with the data from the 36 hands in the CTS group and 15 hands in the non-CTS group that were not used for the training. The reconstruction error of the AE was calculated using the mean square error of the difference between the input and output. By training the AE on non-CTS data only, we could detect patients with CTS because the reconstruction error was smaller for non-CTS data and larger for CTS data. We generated a receiver operating characteristic (ROC) curve by adjusting the cutoff value of the mean square error and calculated the area under the curve (AUC). The optimal cutoff value was set at the point where the Youden index was at its maximum in the ROC curve. Furthermore, to investigate which directional movements contribute to the diagnosis of CTS, we also generated modified screening models that classified CTS and non-CTS using data from only 4 consecutive directions of the 12 directions and calculated the AUC in the same way as above.

Results

The characteristics of the participants are summarized in Table 1. There was no significant difference between the groups in terms of age, sex, or side of the playing hand.

Table 1. Characteristics of participants in the CTS and non-CTS groups.

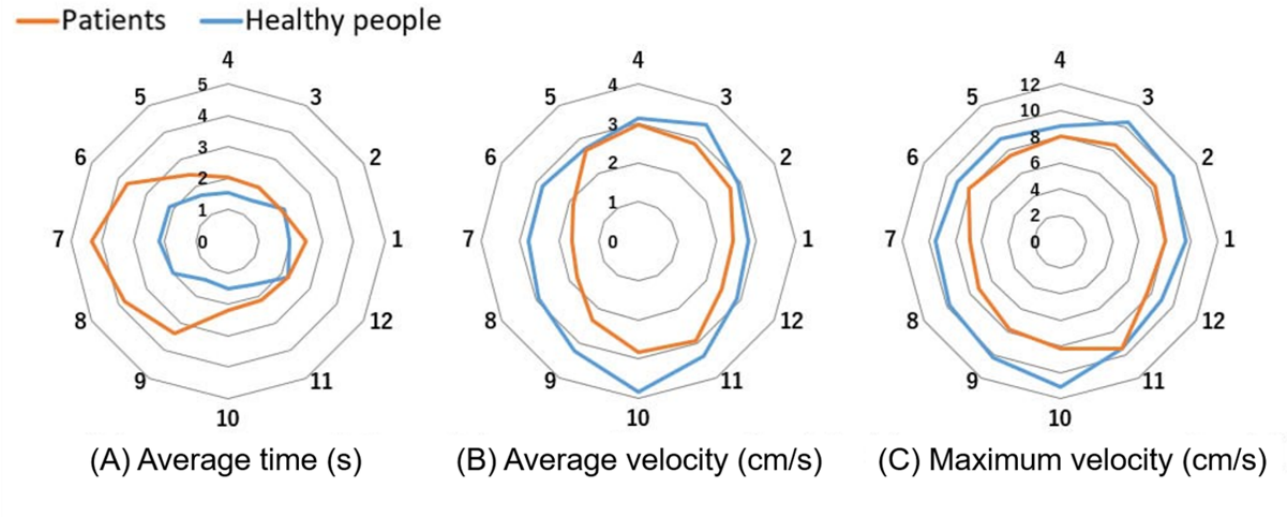
Characteristic	Non-CTS[a]	CTS	*P* value
Participants, n	15	21	N/A[b]
Age (years), mean (SD)	63.5 (17.6)	64.3 (12.2)	.87
Sex (female), n	12	16	.63
Hand dominance (right), n	15	21	>.99
Hands, n	27	36	N/A
Side (right), n	15	17	.69
Bland classification, n			N/A
Grade 1	N/A	5	
Grade 2	N/A	6	
Grade 3	N/A	15	
Grade 4	N/A	0	
Grade 5	N/A	9	
Grade 6	N/A	1	

[a]CTS: carpal tunnel syndrome.

[b]N/A: not applicable.

Figure 6 shows the average time taken to collect vegetables and the average and maximum velocities in each direction. Compared with healthy people, the patients took significantly longer to collect the vegetable in directions 6 to 9, and both the average and maximum velocities of the patients were significantly slower in all directions.

Figure 6. Representation of the average time taken to collect vegetables (A) and the average (B) and maximum (C) velocities in each direction.



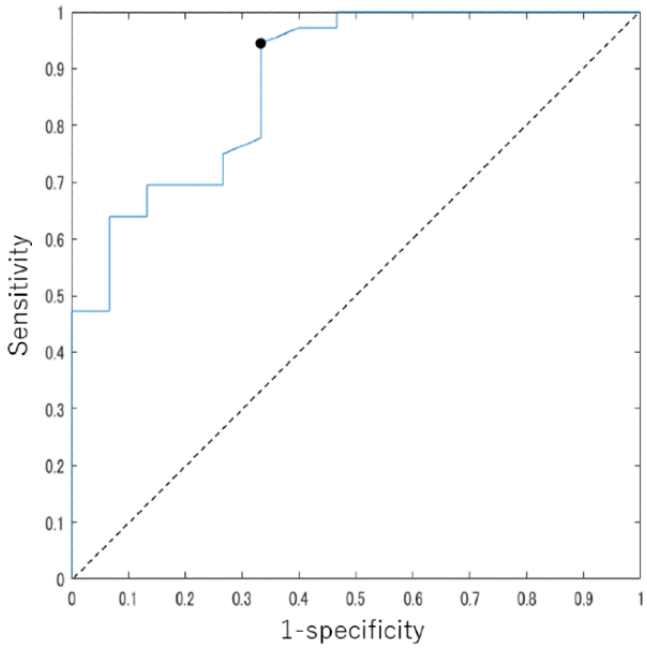
The results of the screening model are shown in Table 2. The participants with and without CTS were classified with 94% sensitivity and 67% specificity. The ROC curve of the classification model is shown in Figure 7; the AUC was 0.86.

Table 2. The result of the screening model. People with and without CTS were classified with 94% sensitivity and 67% specificity.

True label	Predicted label, n	
	Non-CTS[a]	CTS
Non-CTS	10	5
CTS	2	34

[a]CTS: carpal tunnel syndrome.

Figure 7. ROC curve of the screening model. The area under the ROC curve was 0.86. The black point indicates the optimal cutoff value, and the sensitivity and specificity at that point were 0.94 and 0.67, respectively. ROC: receiver operating characteristic.



The results of the modified screening models are shown in Table 3. The model using data from directions 8 to 11 had the highest AUC of 0.99 and could classify patients as CTS and non-CTS with 92% sensitivity and 100% specificity.

Table 3. The index of the modified screening models.

Direction[a]	Sensitivity, %	Specificity, %	AUC[b]
1-4	78	73	0.85
2-5	89	93	0.96
3-6	83	80	0.87
4-7	100	87	0.92
5-8	94	73	0.86
6-9	89	80	0.86
7-10	86	87	0.92
8-11	92	100	0.99
9-12	92	100	0.98
10-1	92	87	0.94
11-2	92	80	0.86
12-3	73	73	0.79

[a]Directions are based on screens B and C of Figure 2.

[b]AUC: area under the curve.

Discussion

Principal Results

In this study, we developed a smartphone app with a high ability to screen for CTS. The app could diagnose CTS with 94% sensitivity and 67% specificity and was almost equal to a tablet app in a previous study, which diagnosed CTS with 93% sensitivity and 73% specificity [19]. The result was also as good as physical examinations; the Tinel sign showed 23% to 60% sensitivity and 64% to 87% specificity, and the Phalen test showed 51% to 91% sensitivity and 33% to 86% specificity in previous studies [11,12]. As we could obtain the same diagnostic ability as physical examinations without a direct medical examination, the app would be useful for screening for CTS in telemedicine in the circumstances of COVID-19.

In the modified screening models, the model using data from directions 8 to 11 had the highest AUC of 0.99 and could diagnose CTS with 92% sensitivity and 100% specificity; this was better than the screening model that used data in all directions. This result suggests that thumb movement from

directions 8 to 11 is different between the CTS and non-CTS groups, contributing to the diagnosis of CTS. Reaching directions 8 to 11 requires a movement similar to thumb opposition, as in screen B of Figure 3, a movement that is impaired in people with CTS [3]. This difficulty with thumb opposition was apparent when using our system.

We used a similar app as in the previous study [19] but with 2 novel aspects. First, our system used a smartphone instead of a tablet. The usage rate of smartphones in Japan is approximately 80% in people aged 40 years or older, higher than that of tablets (40%) [21]. Since we intend to use this app as a screening tool for CTS, it should be accessible to many people. Therefore, it is important to use common equipment. Smartphones have been used in many medical studies because of their utility and universality [16]. Second, our system used anomaly detection algorithms (instead of a binary classification), which have been studied extensively in the detection of system failures in infrastructure and factories, malware detection, and computer vision [22]. Anomaly detection algorithms are also used in medicine, such as medical images [25,26], electrocardiograms [27], and remote medicine [28,29]. Although classification techniques are the most common approaches to anomaly detection, data sets often lack sufficient labeled anomalies. In such cases, unsupervised anomaly detection using statistical and machine learning is more promising. The binary classification used in the previous study requires 2 data sets, one from healthy people and one from patients. In contrast, anomaly detection algorithms require only data sets of healthy people. In general, large data sets are required to enhance machine learning. If our app is used widely, it will be easier to collect data sets from healthy people than patients with CTS. Thus, our system can be enhanced easily in the future.

Limitations

This study has some limitations. First, the varied sizes of the participants' hands were not considered. Healthy people with small hands who struggled to reach each direction may have been misdiagnosed with CTS. Second, because smartphone sizes vary, the level of difficulty depends on the model. Therefore, it is desirable to adjust the size of the circle in the game before playing according to the size of each player's fingers and the smartphone. Third, we used an inexpensive finger guide on the back of the smartphone to fix the hand. If special equipment is required, few people will be able to use our system. It would be better to use readily available equipment, such as fall prevention devices for smartphones, instead. Fourth, while we obtained good results in this study, there is still room for further improvement in machine learning. In order to take advantage of anomaly detection, it is desirable to collect more samples. Finally, our system diagnosed only the presence of CTS. In future work, we will improve our system by collecting more data sets to enable estimation of the severity of CTS.

Conclusions

We developed an app for screening patients with CTS that revealed the difficulty of thumb opposition for patients with CTS and could screen for CTS with high sensitivity and specificity. The app can be used by many people because it is smartphone based, and the machine learning is easy to enhance using anomaly detection. In future work, we will enhance our system by collecting more data sets to enable estimation of the severity of CTS.

Acknowledgments

This research was supported by JST AIP-PRISM JPMJCR18Y2 and JST PRESTO JPMJPR17J4. The authors thank the staff and patients at Sajima Osteopathic Hospital.

Conflicts of Interest

None declared.

References

1. Bongers F, Schellevis F, van den Bosch WJHM, van der Zee J. Carpal tunnel syndrome in general practice (1987 and 2001): incidence and the role of occupational and non-occupational factors. Br J Gen Pract 2007 Jan;57(534):36-39 [FREE Full text] [Medline: 17244422]
2. Padua L, Coraci D, Erra C, Pazzaglia C, Paolasso I, Loreti C, et al. Carpal tunnel syndrome: clinical features, diagnosis, and management. Lancet Neurol 2016 Nov;15(12):1273-1284. [doi: 10.1016/S1474-4422(16)30231-9] [Medline: 27751557]
3. Kuroiwa T, Fujita K, Nimura A, Miyamoto T, Sasaki T, Okawa A. A new method of measuring the thumb pronation and palmar abduction angles during opposition movement using a three-axis gyroscope. J Orthop Surg Res 2018 Nov 16;13(1):288 [FREE Full text] [doi: 10.1186/s13018-018-0999-3] [Medline: 30445972]
4. Wipperman J, Goerl K. Carpal Tunnel Syndrome: Diagnosis and Management. Am Fam Physician 2016 Dec 15;94(12):993-999 [FREE Full text] [Medline: 28075090]
5. Atroshi I, Gummesson C, Johnsson R, Ornstein E, Ranstam J, Rosén I. Prevalence of carpal tunnel syndrome in a general population. JAMA 1999 Jul 14;282(2):153-158. [doi: 10.1001/jama.282.2.153] [Medline: 10411196]
6. Bland JDP, Rudolfer SM. Clinical surveillance of carpal tunnel syndrome in two areas of the United Kingdom, 1991-2001. J Neurol Neurosurg Psychiatry 2003 Dec;74(12):1674-1679 [FREE Full text] [doi: 10.1136/jnnp.74.12.1674] [Medline: 14638888]

7. Burke DT, Burke MM, Stewart GW, Cambré A. Splinting for carpal tunnel syndrome: in search of the optimal angle. Arch Phys Med Rehabil 1994 Nov;75(11):1241-1244. [doi: 10.1016/0003-9993(94)90012-4] [Medline: 7979936]
8. Marshall S, Tardif G, Ashworth N. Local corticosteroid injection for carpal tunnel syndrome. Cochrane Database Syst Rev 2007 Apr 18;2:CD001554. [doi: 10.1002/14651858.CD001554.pub2] [Medline: 17443508]
9. Wintman BI, Winters SC, Gelberman RH, Katz JN. Carpal tunnel release. Correlations with preoperative symptomatology. Clin Orthop Relat Res 1996 May;326:135-145. [Medline: 8620634]
10. Gong HS, Oh JH, Bin SW, Kim WS, Chung MS, Baek GH. Clinical features influencing the patient-based outcome after carpal tunnel release. J Hand Surg Am 2008 Nov;33(9):1512-1517. [doi: 10.1016/j.jhsa.2008.05.020] [Medline: 18984332]
11. D'Arcy CA, McGee S. The rational clinical examination. Does this patient have carpal tunnel syndrome? JAMA 2000 Jun 21;283(23):3110-3117. [doi: 10.1001/jama.283.23.3110] [Medline: 10865306]
12. MacDermid JC, Wessel J. Clinical diagnosis of carpal tunnel syndrome: a systematic review. J Hand Ther 2004;17(2):309-319. [doi: 10.1197/j.jht.2004.02.015] [Medline: 15162113]
13. Basiri K, Katirji B. Practical approach to electrodiagnosis of the carpal tunnel syndrome: A review. Adv Biomed Res 2015;4:50 [FREE Full text] [doi: 10.4103/2277-9175.151552] [Medline: 25802819]
14. Fowler JR, Cipolli W, Hanson T. A Comparison of Three Diagnostic Tests for Carpal Tunnel Syndrome Using Latent Class Analysis. J Bone Joint Surg Am 2015 Dec 02;97(23):1958-1961. [doi: 10.2106/JBJS.O.00476] [Medline: 26631997]
15. Sonoo M, Menkes DL, Bland JD, Burke D. Nerve conduction studies and EMG in carpal tunnel syndrome: Do they add value? Clin Neurophysiol Pract 2018;3:78-88 [FREE Full text] [doi: 10.1016/j.cnp.2018.02.005] [Medline: 30215013]
16. Ienaga N, Fujita K, Koyama T, Sasaki T, Sugiura Y, Saito H. Development and User Evaluation of a Smartphone-Based System to Assess Range of Motion of Wrist Joint. J Hand Surg Glob Online 2020 Nov;2(6):339-342 [FREE Full text] [doi: 10.1016/j.jhsg.2020.09.004] [Medline: 33083772]
17. Asselborn T, Gargot T, Kidziński Ł, Johal W, Cohen D, Jolly C, et al. Automated human-level diagnosis of dysgraphia using a consumer tablet. NPJ Digit Med 2018;1:42 [FREE Full text] [doi: 10.1038/s41746-018-0049-x] [Medline: 31304322]
18. Masaki K, Tateno H, Nomura A, Muto T, Suzuki S, Satake K, et al. A randomized controlled trial of a smoking cessation smartphone application with a carbon monoxide checker. NPJ Digit Med 2020;3:35 [FREE Full text] [doi: 10.1038/s41746-020-0243-5] [Medline: 32195370]
19. Fujita K, Watanabe T, Kuroiwa T, Sasaki T, Nimura A, Sugiura Y. A Tablet-Based App for Carpal Tunnel Syndrome Screening: Diagnostic Case-Control Study. JMIR Mhealth Uhealth 2019 Sep 13;7(9):e14172 [FREE Full text] [doi: 10.2196/14172] [Medline: 31586365]
20. Ministry of Internal Affairs and Communications. Inspection report about time of using media and activity for information in 2020 [Japanese]. Japan Ministry of Internal Affairs and Communications. URL: https://www.soumu.go.jp/main_content/000708016.pdf [accessed 2020-12-11]
21. Mobile fact sheet. Pew Research Center. 2019 Jun 12. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2020-12-11]
22. Chandola V, Banerjee A, Kumar V. Anomaly detection. ACM Comput Surv 2009 Jul;41(3):1-58. [doi: 10.1145/1541880.1541882]
23. Bland JD. A neurophysiological grading scale for carpal tunnel syndrome. Muscle Nerve 2000 Aug;23(8):1280-1283. [doi: 10.1002/1097-4598(200008)23:8<1280::aid-mus20>3.0.co;2-y] [Medline: 10918269]
24. Chalapathy R, Chawla S. Deep Learning for Anomaly Detection: A Survey. ArXiv. Preprint posted online January 23, 2019 2021 [FREE Full text]
25. Wei Q, Ren Y, Hou R, Shi B, Lo JY, Carin L. Anomaly detection for medical images based on a one-class classification. 2018 Presented at: SPIE Medical Imaging 2018: Computer-Aided Diagnosis; Feb 10-15, 2018; Houston, TX. [doi: 10.1117/12.2293408]
26. Schlegl T, Seeböck P, Waldstein SM, Langs G, Schmidt-Erfurth U. f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. Med Image Anal 2019 May;54:30-44. [doi: 10.1016/j.media.2019.01.010] [Medline: 30831356]
27. Zhang CK, Chen YY, Yin A, Wang X. Anomaly detection in ECG based on trend symbolic aggregate approximation. Math Biosci Eng 2019 Mar 12;16(4):2154-2167 [FREE Full text] [doi: 10.3934/mbe.2019105] [Medline: 31137204]
28. Adler DA, Ben-Zeev D, Tseng VW, Kane JM, Brian R, Campbell AT, 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]
29. 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 Nov 13;8(11):e21209 [FREE Full text] [doi: 10.2196/21209] [Medline: 33185562]

Abbreviations

AE: autoencoder
AUC: area under the curve

CTS: carpal tunnel syndrome
NCS: nerve conduction study
ROC: receiver operating characteristic

Edited by G Eysenbach; submitted 11.12.20; peer-reviewed by S Kriventsov, A Oetomo, N Jalali; comments to author 05.01.21; revised version received 15.02.21; accepted 27.02.21; published 14.03.21.

Please cite as:
Koyama T, Sato S, Toriumi M, Watanabe T, Nimura A, Okawa A, Sugiura Y, Fujita K
A Screening Method Using Anomaly Detection on a Smartphone for Patients With Carpal Tunnel Syndrome: Diagnostic Case-Control Study
JMIR Mhealth Uhealth 2021;9(3):e26320
URL: https://mhealth.jmir.org/2021/3/e26320
doi:10.2196/26320
PMID:33714936

Original Paper

Text Messaging and Web-Based Survey System to Recruit Patients With Low Back Pain and Collect Outcomes in the Emergency Department: Observational Study

Anita Barros Amorim[1,2], PhD; Danielle Coombs[2,3], BAppSc; Bethan Richards[2,4], PhD; Chris G Maher[2], FAAHMS; Gustavo C Machado[2], PhD

[1]Discipline of Physiotherapy, Sydney School of Health Sciences, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia
[2]Institute for Musculoskeletal Health, The University of Sydney and Sydney Local Health District, Sydney, Australia
[3]Physiotherapy Department, Royal Prince Alfred Hospital, Sydney, Australia
[4]Rheumatology Department, Royal Prince Alfred Hospital, Sydney, Australia

Corresponding Author:
Anita Barros Amorim, PhD
Discipline of Physiotherapy, Sydney School of Health Sciences
Faculty of Medicine and Health
The University of Sydney
Level 7 East, Susan Wakil Health Building D18
Camperdown
Sydney, 2006
Australia
Phone: 61 0401399572
Email: anita.amorim@sydney.edu.au

Abstract

Background: Low back pain (LBP) is a frequent reason for emergency department (ED) presentations, with a global prevalence of 4.4%. Despite being common, the number of clinical trials investigating LBP in the ED is low. Recruitment of patients in EDs can be challenging because of the fast-paced and demanding ED environment.

Objective: The aim of this study is to describe the recruitment and response rates using an SMS text messaging and web-based survey system supplemented by telephone calls to recruit patients with LBP and collect health outcomes in the ED.

Methods: An automated SMS text messaging system was integrated into Research Electronic Data Capture and used to collect patient-reported outcomes for an implementation trial in Sydney, Australia. We invited patients with nonserious LBP who presented to participating EDs at 1, 2, and 4 weeks after ED discharge. Patients who did not respond to the initial SMS text message invitation were sent a reminder SMS text message or contacted via telephone. The recruitment rate was measured as the proportion of patients who agreed to participate, and the response rate was measured as the proportion of participants completing the follow-up surveys at weeks 2 and 4. Regression analyses were used to explore factors associated with response rates.

Results: In total, 807 patients with nonserious LBP were invited to participate and 425 (53.0%) agreed to participate. The week 1 survey was completed by 51.5% (416/807) of participants. At week 2, the response rate was 86.5% (360/416), and at week 4, it was 84.4% (351/416). Overall, 60% of the surveys were completed via SMS text messaging and on the web and 40% were completed via telephone. Younger participants and those from less socioeconomically disadvantaged areas were more likely to respond to the survey via the SMS text messaging and web-based system.

Conclusions: Using an SMS text messaging and web-based survey system supplemented by telephone calls is a viable method for recruiting patients with LBP and collecting health outcomes in the ED. This hybrid system could potentially reduce the costs of using traditional recruitment and data collection methods (eg, face-to-face, telephone calls only).

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2017-019052

(JMIR Mhealth Uhealth 2021;9(3):e22732) doi:10.2196/22732

KEYWORDS

emergency department; clinical trial; low back pain; acute pain; data collection; patient recruitment; short message service; patient reported outcome measures; mobile phone

Introduction

Background

Low back pain (LBP) is a frequent reason for emergency department (ED) presentations, with a global prevalence of 4.4% [1]. This places LBP in the top 10 presenting complaints in the ED [2], as a broad spectrum of illnesses and injuries are seen in this setting. For comparison, the most common reason for visits to EDs (*injuries and adverse effects*) has a prevalence of 18% and the second most common reason (*cough, upper respiratory symptoms, or ears/nose/throat symptoms*) has a prevalence of 9% [3].

Despite being common, the number of clinical trials investigating LBP in emergency settings is surprisingly low [4]. One possible reason is the challenge in recruiting and collecting patient-reported outcomes in such a busy environment—EDs are often overcrowded, patients might present after hours, emergency clinicians do not have sufficient time, and there is a lack of administrative and structural support for research [5]. Employing effective and efficient methods for recruiting study participants and collecting data is crucial for optimizing research in this setting.

Mobile phones and internet-based technologies have been widely used to recruit and collect data for clinical trials in recent times [6,7]. Advantages regarding accessibility such as being low cost and time efficient and providing access to hard-to-reach populations make this method appealing for research [8]. There is also evidence showing that data collected via mobile phones and internet-based technologies are reliable, valid, and feasible [9,10]. For example, the use of mobile phone services, such as SMS text messaging, has been tested for data collection in LBP research in primary care, showing high response rates [11]. However, little is known about the use of SMS text messaging in combination with a web-based survey system to recruit participants and collect health outcomes in ED settings.

Objectives

To explore the use of contemporary, low-cost, and more efficient options for conducting clinical research in EDs, we investigated the recruitment and response rates when using an SMS text messaging and web-based survey system, supplemented by telephone interviews, within a stepped-wedge cluster randomized controlled trial [12]. The primary aim of this study is to investigate the recruitment and response rates after an ED presentation for LBP. As previous research has suggested that participants' characteristics, such as age, sex, and socioeconomic status, can significantly influence response rates [13], we also aim to investigate whether these factors influenced response rates in the ED setting.

Methods

Design

This is an observational study nested within a stepped-wedge cluster randomized controlled trial that evaluated the implementation of an evidence-based model of care for LBP in four public hospital EDs. The protocol for the Sydney Health Partners Emergency Department (SHaPED) trial has been published elsewhere [12]. The study design, procedures, and informed consent were approved by the Human Research Committees of the Sydney Local Health District (Royal Prince Alfred Hospital zone, protocol number X17-0043).

Participants

Participants were recruited between July and December 2018 from the EDs of four public hospitals in New South Wales (NSW), Australia: Concord Repatriation General Hospital, Royal Prince Alfred (RPA) Hospital, Canterbury Hospital, and Dubbo Base Hospital. Eligible patients were identified using discharge diagnosis codes from the Systematized Nomenclature of Medicine—Clinical Terms—Australian version, Emergency Department Reference Set [14] (Multimedia Appendix 1). Only patients presenting to the EDs with nonserious forms of LBP (nonspecific LBP, sciatica, and lumbar spinal stenosis) and with a mobile phone number recorded in the medical records were invited to participate. Representations to the ED within 48 hours or LBP related to serious spinal pathologies, such as lumbar fracture, infection, malignancy, or cauda equina syndrome, were excluded.

Recruitment

Upon discharge from the ED, eligible patients were informed about the study by clinical staff and/or received a flyer with information about a text message invitation and web-based survey. This method was used to ensure that the patients were aware of the survey before receiving the SMS text message invitation. Then, local clinical staff obtained patients' information (ie, name, mobile number, and postcode) from the hospital's electronic medical records. Patients' information was inserted into a secure web app (Research Electronic Data Capture [REDCap]) by research staff.

An automated SMS text messaging system (Twilio Inc) was integrated into REDCap and used to schedule SMS text message invitations. We used an opt-out approach to ensure that there was no pressure or coercion on patients to consent to participate in the survey. Seven days after the ED visit at 12:30 PM, an SMS text message was sent to eligible patients with an invitation and link to answer the web-based survey. Patients who did not want to be invited to participate in the study could inform the local clinical staff at the time of discharge, who would then notify researchers to remove them from the invitation list. Potential participants could also ignore the SMS text message invitation or opt out by replying *NO* and would no longer be contacted. REDCap was scheduled to send an SMS text message

(via Twilio) containing the following text approved by the Human Research Ethics Committee: “Dear [name], hope you are going well after your recent visit for back pain to our ED. We are interested in how your back is going and what you thought of our care. Our survey will take only 5 minutes to complete. This survey is being conducted by Dr [name], ED Director of [hospital]. To opt out reply NO -- to begin the survey, visit [link]”

The link in the SMS text message invitation referred eligible patients to the web-based participant information statement and consent information. At this point, potential participants could decline to participate and would no longer be contacted. Those who agreed to participate were referred to a brief self-reported web-based survey aimed at collecting patient-reported outcomes (Multimedia Appendix 2). Completion of the web-based survey indicated consent to participate in the survey. The survey was not anonymous, and patients did not receive any financial remuneration for responding to it.

For those who agreed to participate and completed the week 1 survey, 2 follow-up surveys were sent at 2 and 4 weeks after ED presentation. Initially, we scheduled reminder messages to be sent 3 times (on consecutive days) for each data collection wave if the survey had not been completed or the patient had not opted out of the study. Therefore, each participant had up to four opportunities to respond to the web-based survey directly on their smartphone. Halfway through the study, we changed the scheduling system to one reminder only, as suggested by our Human Research Ethics Committee, to avoid patients becoming overwhelmed by the number of text message invitations. To maximize the response rate, participants who did not respond to the final reminder message were contacted verbally via telephone. Not completing the week 2 survey, either on the web or via telephone, did not preclude participants from completing the week 4 survey. The research staff followed a script approved by the Human Research Ethics Committee to conduct the telephone survey.

Predictors and Outcomes

In this study, we investigated recruitment and response rates as outcomes. Patients’ responses to 3 surveys (ie, at 1, 2, and 4 weeks after discharge) were classified as *yes* (survey completed) or *no* (survey not completed or the patient opted out of the study). We also classified (yes/no) whether responses occurred at the initial SMS text message invitation, after reminder messages, or during telephone calls. Recruitment rates were measured as the proportion of eligible patients consenting to participate upon first SMS text message invitation, after a reminder SMS text message, or during telephone calls (yes/no). Response rates were measured as the proportion of included participants completing the follow-up surveys at weeks 2 and 4 via initial SMS text message invitation, reminder SMS text message, or telephone calls (yes/no).

We had limited access to patient’s information and extracted the following putative predictors from the hospital’s electronic medical records: age, sex, and postcode. The Australian Bureau of Statistics’ Socio-Economic Indexes for Areas (SEIFA) 2016 [15] was used to classify patients’ socioeconomic status based on their postcode of residence. SEIFA was reported as deciles, with the lowest decile designating areas with the greatest socioeconomic disadvantage.

Analyses

Descriptive analyses were used to report recruitment and response rates (via initial SMS text message, reminder SMS text message, or telephone calls) for all participants and grouped by recruitment site, age, sex, and socioeconomic status. Recruitment and response rates were also calculated for each month during the 6-month trial period. A multiple logistic regression model was used to evaluate whether age, sex, and socioeconomic status (SEIFA deciles) were predictors of whether a participant responded to the survey on the web or via telephone calls. In contrast to the other 3 metropolitan Sydney sites included in the parent study, Dubbo Base Hospital is located in a rural area of NSW. Thus, we decided to analyze whether there were any differences in recruitment or response rates between the metropolitan and rural EDs. All data analyses were conducted using STATA (version 14.0, STATA Corporation).

Results

Recruitment Rate

Data were collected between July and December 2018. In total, 807 eligible patients with nonserious LBP were invited to participate and 425 (53.0%) agreed to participate. After 9 participants dropped out without reasons, 51.5% (416/807) entered the study and completed the week 1 survey. At week 2, 86.5% (360/416) participants completed the follow-up survey, and at week 4, 84.4% (351/416) completed the survey (Figure 1). The highest recruitment rates were in females (212/392, 54.1%), at Concord Hospital (113/195, 57.9%), among people aged 40 to 69 years (83/141, 58.9% - 68/109, 62.4%), and from people living in the least socioeconomic disadvantaged areas (61/107, 57.0% - 35/58, 60.3%—SEIFA deciles 8, 9, and 10; Table 1). Overall, 59.6% (248/416) of participants were recruited via SMS text messaging alone (including the reminder SMS text message) and 40.4% (168/416) agreed to participate via telephone calls.

Figure 1. Study flowchart.

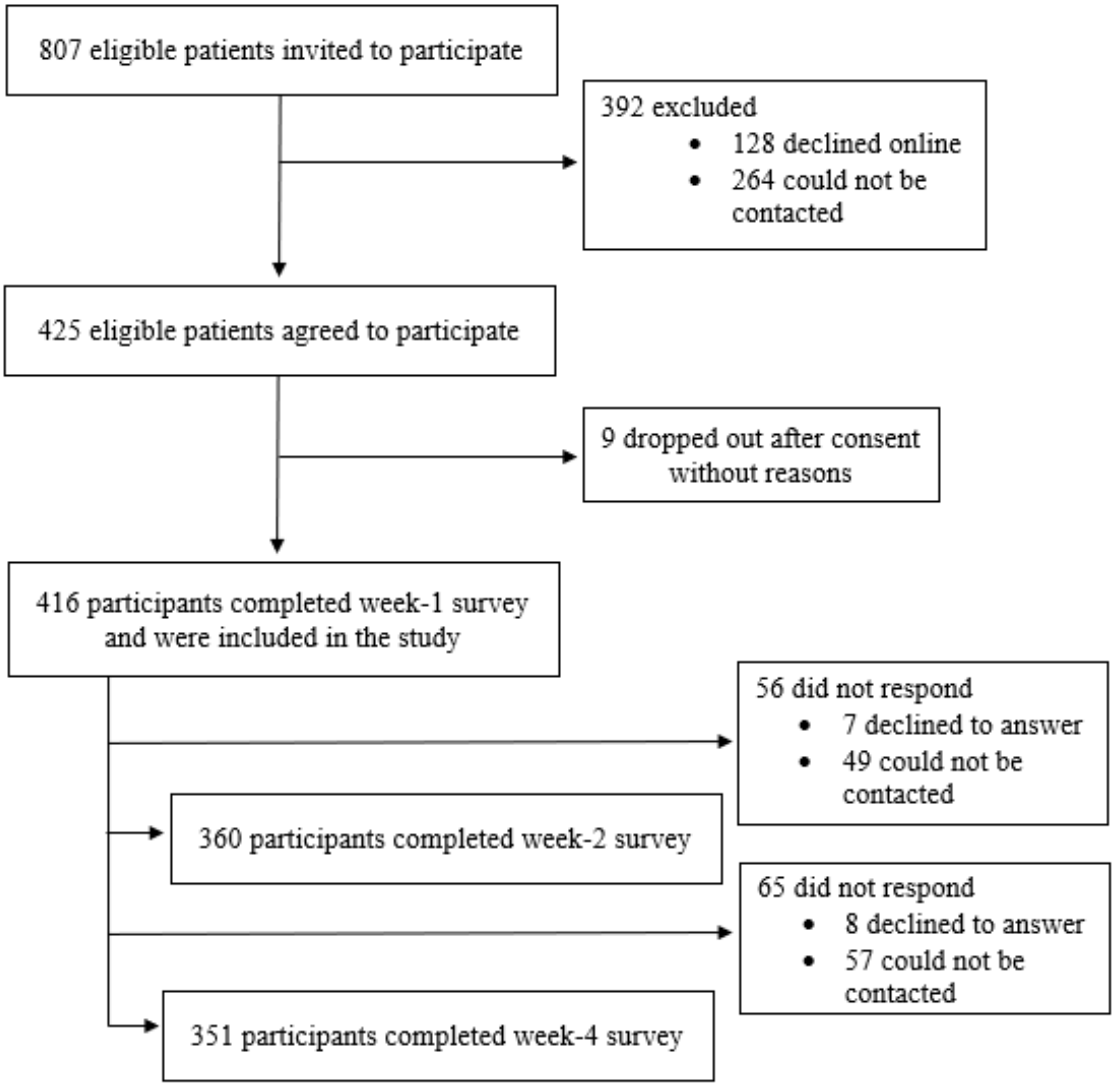


Table 1. Recruitment rates by response method, grouped by recruitment site, age, sex, and socioeconomic status (N=807).

Variable	Invited to participate	Recruited via initial SMS text message	Recruited via reminder SMS text message	Recruited via telephone calls	Total recruited participants
Participant, n (%)	807 (100)	138 (17.1)	110 (13.6)	168 (20.8)	416 (51.5)
Recruitment site, n (%)					
RPA[a]	292 (36.1)	70 (23.9)	42 (14.4)	46 (15.8)	158 (54.1)
Canterbury	196 (24.3)	24 (12.2)	24 (12.2)	47 (23.9)	95 (48.5)
Concord	195 (24.2)	35 (17.9)	27 (13.8)	51 (26.2)	113 (57.9)
Dubbo	124 (15.4)	9 (7.3)	17 (13.6)	24 (19.2)	50 (40.3)
Age group (years), n (%)					
18-29	123 (15.2)	14 (11.4)	17 (13.8)	18 (14.6)	49 (39.8)
30-39	158 (19.6)	29 (18.3)	19 (12.0)	28 (17.7)	76 (48.1)
40-49	141 (17.5)	26 (18.4)	27 (19.1)	30 (21.3)	83 (58.9)
50-59	134 (16.6)	27 (20.1)	22 (16.4)	30 (22.4)	79 (59.0)
60-69	109 (13.5)	26 (23.8)	10 (9.2)	32 (29.4)	68 (62.4)
70-79	84 (10.4)	11 (13.1)	6 (7.1)	21 (25.0)	38 (45.2)
80+	57 (7.1)	5 (8.8)	9 (15.8)	9 (15.8)	23 (40.0)
Sex, n (%)					
Female	392 (48.6)	72 (18.4)	62 (15.8)	78 (19.9)	212 (54.1)
Male	415 (51.4)	66 (15.9)	48 (11.6)	90 (21.7)	204 (49.2)
Socioeconomic status (SEIFA[b] deciles), n (%)					
1[c]	41 (5.1)	9 (21.9)	3 (7.3)	7 (17.1)	19 (46.3)
2	39 (4.8)	5 (12.8)	6 (15.4)	10 (25.6)	21 (53.8)
3	79 (9.8)	8 (10.1)	11 (13.9)	22 (27.8)	41 (51.9)
4	124 (15.4)	12 (9.7)	16 (12.9)	23 (18.5)	51 (41.1)
5	4 (0.5)	0 (0.0)	2 (50.0)	0 (0.0)	2 (50.0)
6	68 (8.4)	13 (19.1)	10 (14.7)	10 (14.7)	33 (48.5)
7	81 (10.0)	11 (13.5)	10 (12.3)	18 (22.2)	39 (48.1)
8	58 (7.2)	13 (22.4)	10 (17.2)	12 (20.7)	35 (60.3)
9	107 (13.3)	24 (22.4)	13 (12.1)	24 (22.4)	61 (57.0)
10	206 (25.5)	43 (20.9)	29 (14.1)	42 (20.4)	114 (55.3)

[a]RPA: Royal Prince Alfred.

[b]SEIFA: Socio-Economic Indexes for Areas.

[c]Decile 1 contains the most disadvantaged areas.

Response Rate

Table 2 shows the characteristics of the study participants (n=416) and the response rates at weeks 2 and 4. Participants' mean age was 50.0 years (SD 17.1), 51% were female, most participants (38%) presented at the RPA Hospital ED, and half (50%) were from the least socioeconomic disadvantaged areas (SEIFA deciles 8, 9, and 10). The highest response rates at week 2 were at RPA Hospital (90%), among people aged 50 to 59 years (96%), females (87%), and those from the least socioeconomic disadvantaged areas (SEIFA deciles 8 [100%] and 10 [93%]). In contrast, response rates at week 4 were higher among older people (aged ≥80 years, 96%) and males (88%); however, similar rates were found in the other categories.

Table 2. Participants' characteristics and response rates at weeks 2 and 4, grouped by recruitment site, age, sex, and socioeconomic status (n=416).

Variables	Total[a]	Responded at week 2[b]	Responded at week 4[b]
Participant, n (%)	416 (100)	360 (86.5)	351 (84.4)
Recruitment site, n (%)			
RPA[c]	158 (38.0)	142 (89.9)	142 (89.9)
Canterbury	95 (22.8)	80 (84.2)	76 (80.0)
Concord	113 (27.2)	96 (85.0)	88 (77.9)
Dubbo	50 (12.0)	42 (84.0)	45 (90.0)
Age group (years), n (%)			
18-29	49 (11.7)	35 (71.4)	33 (67.3)
30-39	76 (18.3)	64 (84.2)	62 (81.6)
40-49	83 (19.9)	69 (83.1)	68 (81.9)
50-59	79 (19.0)	76 (96.2)	71 (89.9)
60-69	68 (16.4)	61 (89.7)	63 (92.6)
70-79	38 (9.1)	34 (89.5)	32 (84.2)
80+	23 (5.6)	21 (91.3)	22 (95.6)
Sex, n (%)			
Female	212 (51.0)	185 (87.3)	171 (80.7)
Male	204 (49.0)	175 (85.8)	180 (88.2)
Socioeconomic status (SEIFA[d] deciles), n (%)			
1[e]	19 (4.6)	17 (89.5)	14 (73.7)
2	21 (5.0)	17 (80.9)	18 (85.7)
3	41 (9.9)	34 (82.9)	34 (82.9)
4	51 (12.3)	42 (82.3)	44 (86.2)
5	2 (0.5)	1 (50.0)	1 (50.0)
6	33 (7.9)	30 (90.9)	28 (84.8)
7	39 (9.4)	33 (84.6)	31 (79.5)
8	35 (8.4)	35 (100.0)	34 (97.1)
9	61 (14.5)	45 (73.8)	51 (83.6)
10	114 (27.4)	106 (93.0)	96 (84.2)

[a]Study sample characteristics at baseline. Percentages correspond to the number of participants in each group divided by the total number of participants.

[b]Response rates at weeks 2 and 4 grouped by recruitment site, age, sex, and socioeconomic status.

[c]RPA: Royal Prince Alfred.

[d]SEIFA: Socio-Economic Indexes for Areas.

[e]Decile 1 contains the most disadvantaged areas.

Recruitment and Response Rates by Response Method

Table 3 shows recruitment and response rates by each study period and method of response (SMS text messaging and web-based system or telephone calls). Overall, recruitment rates varied during the study from 43% (December 2018) to 61% (September 2018). Response rates via the SMS text messaging and web-based system decreased over time (from 73% to 51%), whereas those via telephone calls increased from 27% to 49% during the study period. The results of a multiple logistic regression model presented in Table 4 show that web-based responses were significantly higher in those who were younger (odds ratio [OR] 0.99, 95% CI 0.97-0.99; *P*=.02) and those in less socioeconomic disadvantaged areas (OR 1.08, 95% CI 1.00-1.16; *P*=.03).

Table 3. Recruitment and response rates by response method, grouped by study period.

Study period (month)	Invited to participate, n	Recruited participants[a], n (%)	Responded via SMS text message (initial and reminder)[b], n (%)	Responded via telephone[b], n (%)
July 2018	147	70 (47.6)	51 (72.9)	19 (27.1)
August 2018	146	74 (50.7)	56 (75.7)	18 (24.3)
September 2018	116	71 (61.2)	35 (49.3)	36 (50.7)
October 2018	154	82 (53.2)	45 (54.9)	37 (45.1)
November 2018	157	82 (52.2)	42 (51.2)	40 (48.8)
December 2018	87	37 (42.5)	19 (51.4)	18 (48.6)
Total	807	416 (51.5)	248 (59.6)	168 (40.4)

[a]Percentages based on the number of people invited to participate in the study.

[b]Percentages based on the number of people included in the study.

Table 4. Associations of age, sex, and socioeconomic status with response method (SMS text messaging and web-based or telephone).

Variables	SMS text messaging and web-based system at week 1 (n=416)		SMS text messaging and web-based system at week 2 (n=360)		SMS text messaging and web-based system at week 4 (n=351)	
	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Age	0.99 (0.97-0.99)	.02	0.99 (0.98-1.00)	.17	1.00 (0.99-1.01)	.92
Sex	0.73 (0.49-1.08)	.12	0.79 (0.52-1.22)	.29	0.74 (0.48-1.13)	.16
Socioeconomic status	1.06 (0.99-1.13)	.08	1.08 (1.00-1.16)	.03	1.03 (0.96-1.11)	.40

[a]OR: odds ratio.

Recruitment and Response Rates by Hospital Site

The overall recruitment rate in rural hospitals (50/124, 40.3%) was 13% lower than that in the metropolitan hospitals (366/683, 53.6%). Recruitment rates via SMS text message invitations differed substantially between rural (26/124, 20.9%) and metropolitan (222/683, 32/5%) areas but were similar for telephone calls (24/124, 19.4% and 144/683, 21.1%, respectively). Overall, at week 1, 59.6% (248/416) participants completed the study invitation and survey via the SMS text messaging and web-based system—33.1% (138/416) after the initial SMS text message invitation and 26.4% (110/416) after the reminders—and 40.4% (168/416) completed the survey over the telephone. At week 2, 61.1% (220/360) participants completed the survey via the SMS and web-based system—36.9% (133/360) after initial SMS text message invitation and 24.2% (87/360) after reminders—and 38.8% (140/360) completed the survey over the telephone. At week 4, 55.8% (196/351) participants completed the survey via the SMS text messaging and web-based system—30.5% (107/351) after the initial SMS text message invitation and 25.4% (89/351) after reminders—and 44.2% (155/351) completed the survey over the telephone.

Overall Study Response Rate

The scheduling system changed from 3 reminders to 1 reminder in the first week of October 2018. After the change, the recruitment rate in the study was only 2% smaller, from 53% in July to September 2018 to 51% in October to December 2018. Responses via SMS text messaging (initial or reminder) had an absolute reduction of 13% (from 66% to 53%) at week 1, 7% (from 65% to 58%) at week 2, and 11% (from 61% to 50%) at week 4. The overall response rate (SMS text messaging plus telephone) was approximately 5% greater (84% vs 89%) at week 2 but 3% lower (86% vs 83%) at week 4 after changing the scheduling system to 1 SMS text message reminder (Multimedia Appendix 3).

Discussion

Principal Findings

Recruitment of patients in research, especially in emergency settings, remains a challenge for many studies because of the fast-paced, demanding environment of EDs. This study explored the recruitment and response rates when using an automated SMS text message (Twilio) in combination with a web-based survey system (REDCap) and/or telephone calls to recruit participants and collect LBP outcomes in the ED. Approximately half of all eligible patients were successfully recruited to the trial, and follow-up response rates ranged from 84% to 86%. This suggests that using an SMS text messaging and web-based survey system supplemented by telephone calls is a viable method for recruitment and data collection in this setting. Younger participants and those living in the least socioeconomically disadvantaged areas were more likely to complete the survey on the web rather than via telephone.

Recruitment rates, particularly those via SMS text message invitation, were higher in metropolitan hospitals than in rural sites. These findings are similar to other studies that have shown that older people and those from more socioeconomically disadvantaged areas are less comfortable using smartphones for research purposes [16-18]. The response rates to the SMS text messaging and web-based methods alone at weeks 2 (61%) and 4 (56%) were noticeably lower than the desirable 85% follow-up rate, which was considered adequate for a clinical trial [19]. Consequently, it appears that SMS text messaging and web-based methods alone may not be practical for use as a substitute for traditional methods (eg, face-to-face, telephone calls) of follow-up contact in ED trials. However, we found that adding telephone calls to an automated SMS text messaging and web-based system of data collection increased response rates by up to 86%. Therefore, the use of a hybrid system (SMS text messaging and web-based survey plus telephone follow-ups), in which traditional methods of data collection are only used for those who do not respond to the automated SMS text messaging and web-based system, is a practical option and could potentially result in significant savings in cost and time.

Poor patient recruitment and response are two well-known threats to feasibility in clinical trials [20], especially in EDs because of the demanding and pressurized workplace [21]. These threats can lead to a considerable waste of financial resources or underpowered studies that report on clinically relevant research questions with insufficient statistical power. Therefore, establishing optimal methods to enhance recruitment and response rates in this particular setting is crucial. Although several recent clinical trials have used mobile technology, such as SMS text messaging, to collect outcomes [22], this study is the first to report the recruitment and response rates when using this approach to recruit patients with LBP and collect outcomes in an ED setting.

Our data support the findings of a similar study conducted by Macedo et al [11], demonstrating that SMS text messaging supplemented with telephone follow-ups, but not SMS text messaging alone, provides excellent follow-up response rates for randomized controlled trials with people with LBP. In contrast to our study, Macedo et al [11] investigated response rates in a cohort of patients who had already been included in the study via traditional face-to-face methods. Thus, this study is the first in the LBP field to examine recruitment rates via SMS text messaging as the primary invitation method. We also demonstrated the recruitment and response rates of patients in EDs using this approach, which have not been previously reported. On the basis of our findings, the use of a hybrid system (SMS text messaging and web-based survey plus telephone follow-ups) is a practical option in this setting and could provide considerable reductions in the costs of recruitment and data collection, as the costly traditional methods would only need to be used for approximately 41% of web-based nonresponders.

Considering the barriers involved in recruitment and follow-up assessment in ED trials and the potential advantages of using an SMS text messaging and web-based system to facilitate this, future studies should focus on methods to enhance compliance with these novel technologies. This is particularly important among older people and those from socioeconomically disadvantaged areas, as our results revealed that these factors were associated with lower response rates via web-based systems. In addition, future studies could perform cost-effective analysis comparing automated web-based systems with traditional methods and the benefits in reducing costs and increasing response rates when using a hybrid method. Future research should also explore patients' experiences using this method.

Limitations

Australia is one of the leading users of smartphones, with 89% of the population owning one, and, surprisingly, market growth is being driven by older generations [23]. The same is true for other high-income countries, such as Norway, the Netherlands, and the United Kingdom [24]. However, in low-income and middle-income countries, the scenario is different, with only approximately 45% of the population owning a smartphone, and only a minor proportion of these are owned by older people [25]. Therefore, generalization of our findings may be limited in those countries. Generalization of our results may also be limited to other health jurisdictions, as participants were predominantly recruited from 2 local health districts in NSW, Australia. Although our participants comprised individuals from diverse age groups and socioeconomic areas, we had limited information on other demographic and clinical characteristics. Another limitation is that when participants did not respond to a round of SMS text messaging, they were still contacted via telephone. Although this may have influenced our overall response rate, we analyzed the response rates separately for each method. Furthermore, in the first half of the trial period, participants received up to 3 reminders to complete the survey, which may be the reason for the increased response rate via the SMS text messaging and web-based system compared with telephone calls during this period.

Conclusions

This study demonstrates that an automated SMS text messaging and web-based system in addition to telephone calls, but not SMS text messaging and web-based systems alone, is a viable option for recruiting patients with LBP and collecting outcome data in ED settings in Australia. This hybrid method is likely to facilitate recruitment and data collection in clinical trials in EDs and potentially reduce the costs of using traditional recruitment and data collection methods. However, generalization of our findings may be limited in the countries with a high percentage of smartphone use, such as the United States, Spain, and Germany, where smartphone ownership rates range from 85% to 89%. Future cost-effective analysis should be conducted in similar studies to allow for a clear conclusion regarding the potential cost savings of this method.

Acknowledgments

CM was supported by a National Health and Medical Research Council (NHMRC) Principal Research Fellowship. GM was supported by an NHMRC Early Career Fellowship. The SHaPED trial received seed funding (Aus $110,000 [US $77,500]) from Sydney Health Partners and the NSW Agency for Clinical Innovation. The sponsor of this study was the University of Sydney, NSW 2006, Australia.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Systematized Nomenclature of Medicine—Clinical Terms—Australian version, Emergency Department Reference Set codes related to nonserious low back pain presentations.
[PDF File (Adobe PDF File), 119 KB - mhealth_v9i3e22732_app1.pdf]

Multimedia Appendix 2
Patient survey.
[PDF File (Adobe PDF File), 170 KB - mhealth_v9i3e22732_app2.pdf]

Multimedia Appendix 3
Differences in the recruitment and response rates throughout the study when sending the SMS text message reminder 3 times versus 1 time.
[DOCX File , 15 KB - mhealth_v9i3e22732_app3.docx]

References

1. Edwards J, Hayden J, Asbridge M, Gregoire B, Magee K. Prevalence of low back pain in emergency settings: a systematic review and meta-analysis. BMC Musculoskelet Disord 2017 Apr 4;18(1). [doi: 10.1186/s12891-017-1511-7] [Medline: 28376873]
2. Emergency department care 2016–17: Australian hospital statistics. Australian Institute of Health and Welfare. URL: https://www.aihw.gov.au [accessed 2019-08-11]
3. Kocher KE, Meurer WJ, Fazel R, Scott PA, Krumholz HM, Nallamothu BK. National Trends in Use of Computed Tomography in the Emergency Department. Annals of Emergency Medicine 2011 Nov;58(5):452-462.e3. [doi: 10.1016/j.annemergmed.2011.05.020] [Medline: 21835499]
4. Oliveira CB, Amorim HE, Coombs DM, Richards B, Reedyk M, Maher CG, et al. Emergency department interventions for adult patients with low back pain: a systematic review of randomised controlled trials. Emerg Med J 2020 Oct 09;38(1):59-68. [doi: 10.1136/emermed-2020-209588] [Medline: 33037020]
5. Committee on the Future of Emergency Care in the United States Health System Board on Health Care Services. Hospital-Based Emergency Care: At the Breaking Point. In: Future of Emergency Care Series. Washington, DC: The National Academies; 2006.
6. Leach LS, Butterworth P, Poyser C, Batterham PJ, Farrer LM. Online recruitment: feasibility, cost, and representativeness in a study of postpartum women. J Med Internet Res 2017 Mar 8;19(3):e61 [FREE Full text] [doi: 10.2196/jmir.5745] [Medline: 28274906]
7. Johansen B, Wedderkopp N. Comparison between data obtained through real-time data capture by SMS and a retrospective telephone interview. Chiropr Man Therap 2010 May 26;18(1). [doi: 10.1186/1746-1340-18-10] [Medline: 20500900]
8. Fenner Y, Garland SM, Moore EE, Jayasinghe Y, Fletcher A, Tabrizi SN, et al. Web-based recruiting for health research using a social networking site: an exploratory study. J Med Internet Res 2012 Feb;14(1):e20 [FREE Full text] [doi: 10.2196/jmir.1978] [Medline: 22297093]
9. Hatch EE, Hahn KA, Wise LA, Mikkelsen EM, Kumar R, Fox MP, et al. Evaluation of selection bias in an internet-based study of pregnancy planners. Epidemiology 2016 Jan;27(1):98-104 [FREE Full text] [doi: 10.1097/EDE.0000000000000400] [Medline: 26484423]
10. Andreeva VA, Galan P, Julia C, Castetbon K, Kesse-Guyot E, Hercberg S. Assessment of Response Consistency and Respective Participant Profiles in the Internet-based NutriNet-Sante Cohort. American Journal of Epidemiology 2014 Feb 11;179(7):910-916. [doi: 10.1093/aje/kwt431] [Medline: 24521560]
11. Macedo LG, Maher CG, Latimer J, McAuley JH. Feasibility of Using Short Message Service to Collect Pain Outcomes in a Low Back Pain Clinical Trial. Spine 2012;37(13):1151-1155. [doi: 10.1097/brs.0b013e3182422df0] [Medline: 22146289]
12. Machado GC, Richards B, Needs C, Buchbinder R, Harris IA, Howard K, et al. Implementation of an evidence-based model of care for low back pain in emergency departments: protocol for the Sydney Health Partners Emergency Department (SHaPED) trial. BMJ Open 2018 Apr 19;8(4):e019052. [doi: 10.1136/bmjopen-2017-019052] [Medline: PMC5914780]

13. Bexelius C, Merk H, Sandin S, Ekman A, Nyrén O, Kühlmann-Berenzon S, et al. SMS versus telephone interviews for epidemiological data collection: feasibility study estimating influenza vaccination coverage in the Swedish population. Eur J Epidemiol 2008 Dec 10;24(2):73-81. [doi: 10.1007/s10654-008-9306-7] [Medline: 19082745]
14. Hansen DP, Kemp ML, Mills SR, Mercer MA, Frosdick PA, Lawley MJ. Developing a national emergency department data reference set based on SNOMED CT. Med J Aust 2011 Mar 21;194(4):S8-10. [doi: 10.5694/j.1326-5377.2011.tb02934.x] [Medline: 21401491]
15. Australian Institute of Health and Welfare. Socio-Economic Indexes for Areas (SEIFA). 2033.0.55.001 - Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA). Canberra: Australian Bureau of Statistics; 2016. URL: https://meteor.aihw.gov.au/content/index.phtml/itemId/695778 [accessed 2019-08-11]
16. Vaportzis E, Clausen MG, Gow AJ. Older adults perceptions of technology and barriers to interacting with tablet computers: a focus group study. Front Psychol 2017 Oct 4;8:1687 [FREE Full text] [doi: 10.3389/fpsyg.2017.01687] [Medline: 29071004]
17. Ma D, Tina Du J, Cen Y, Wu P. Exploring the adoption of mobile internet services by socioeconomically disadvantaged people. AJIM 2016 Nov 21;68(6):670-693. [doi: 10.1108/ajim-03-2016-0027]
18. SG J, S T, DW K. Smartphone use and internet literacy of senior citizens. J Assist Technol 2016;10(1):-. [doi: 10.1108/jat-03-2015-0006]
19. Maher CG, Moseley AM, Sherrington C, Elkins MR, Herbert RD. A description of the trials, reviews, and practice guidelines indexed in the PEDro database. Phys Ther 2008 Jul;88(9):1068-1077. [doi: 10.2522/ptj.20080002] [Medline: 18635670]
20. Sully BGO, Julious SA, Nicholl J. A reinvestigation of recruitment to randomised, controlled, multicenter trials: a review of trials funded by two UK funding agencies. Trials 2013;14(1):166. [doi: 10.1186/1745-6215-14-166] [Medline: 23758961]
21. Cofield SS, Conwit R, Barsan W, Quinn J. Recruitment and retention of patients into emergency medicine clinical trials. Acad Emerg Med 2010 Oct;17(10):1104-1112. [doi: 10.1111/j.1553-2712.2010.00866.x] [Medline: 21040112]
22. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013 Jan;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
23. Corbett P, Huggins K, Duryea C. Mobile Consumer Survey 2018. Delloite. 2019. URL: https://www2.deloitte.com/ [accessed 2019-08-15]
24. Drumm JWN, Swiegers M, Davey M. Mobile Consumer Survey 2017. Delloite. 2018. URL: https://www2.deloitte.com/ [accessed 2019-08-15]
25. Taylor K, Silver L. Smartphone Ownership is Growing Rapidly Around the World, but Not Always Equally. Pew Research Center. 2019. URL: https://www.pewresearch.org [accessed 2019-08-15]

Abbreviations

ED: emergency department
LBP: low back pain
NHMRC: National Health and Medical Research Council
NSW: New South Wales
OR: odds ratio
REDCap: Research Electronic Data Capture
SEIFA: Socio-Economic Indexes for Areas
SHaPED: Sydney Health Partners Emergency Department

Edited by G Eysenbach; submitted 22.07.20; peer-reviewed by C Weerth, G Biviá; comments to author 24.10.20; revised version received 16.12.20; accepted 24.12.20; published 04.03.21.

Please cite as:
Amorim AB, Coombs D, Richards B, Maher CG, Machado GC
Text Messaging and Web-Based Survey System to Recruit Patients With Low Back Pain and Collect Outcomes in the Emergency Department: Observational Study
JMIR Mhealth Uhealth 2021;9(3):e22732
URL: https://mhealth.jmir.org/2021/3/e22732
doi:10.2196/22732
PMID:33661125

Original Paper

Motivation to Participate in Precision Health Research and Acceptability of Texting as a Recruitment and Intervention Strategy Among Vietnamese Americans: Qualitative Study

Van Ta Park[1], BA, BS, MPH, PhD; Amber Kim[2], BS; In Hyang Cho[2], BS; Bora Nam[1], MS, PhD; Khue Nguyen[3], BS; Quyen Vuong[3], MSW, MBA; Vyjeyanthi S Periyakoil[4], MD; Y Alicia Hong[5], MAA, PhD

[1]Department of Community Health Systems, School of Nursing, University of California, San Francisco, San Francisco, CA, United States

[2]School of Nursing, San Jose State University, San Jose, CA, United States

[3]International Children Assistance Network, Milpitas, CA, United States

[4]Department of Medicine – Primary Care and Population Health, School of Medicine, Stanford University, Stanford, CA, United States

[5]College of Health and Human Services, George Mason University, Fairfax, VA, United States

Corresponding Author:
Van Ta Park, BA, BS, MPH, PhD
Department of Community Health Systems
School of Nursing
University of California, San Francisco
2 Koret Way, N511S
San Francisco, CA, 94143
United States
Phone: 1 415 514 3318
Email: van.park@ucsf.edu

Abstract

Background: The largest effort undertaken in precision health research is the Precision Medicine Initiative (PMI), also known as the All of Us Research Program, which aims to include 1 million or more participants to be a part of a diverse database that can help revolutionize precision health research studies. Research participation from Asian Americans and Pacific Islanders in precision health research is, however, limited; this includes Vietnamese Americans, especially those with limited English proficiency. PMI engagement efforts with underserved communities, including members of minority populations or individuals who have experienced health disparities such as Vietnamese Americans with limited English proficiency, may help to enrich the diversity of the PMI.

Objective: The aim of this study is to examine the attitudes towards and perceptions of precision health, motivations and barriers to participation in precision health research, and acceptability of SMS text messaging as a recruitment and intervention strategy among underserved Vietnamese Americans.

Methods: A community sample of 37 Vietnamese Americans completed a survey and participated in one of 3 focus groups classified by age (18-30, 31-59, and ≥60 years) on topics related to precision health, participation in precision health research, texting or social media use experience, and insights on how to use text messages for recruitment and intervention. Participants were recruited via community organizations that serve Vietnamese Americans, flyers, word of mouth, and Vietnamese language radio announcements.

Results: Most participants had little knowledge of precision health initially. After brief education, they had positive attitudes toward precision health, although the motivation to participate in precision health research varied by age and prior experience of research participation. The main motivators to participate included the desire for more knowledge and more representation of Vietnamese Americans in research. Participants were open to receiving text messages as part of their research participation and provided specific suggestions on the design and delivery of such messages (eg, simple, in both English and Vietnamese). Examples of barriers included misinterpretation of messages, cost (to send text messages), and preferences for different texting platforms across age groups.

Conclusions: This study represents one of the first formative research studies to recruit underserved Vietnamese Americans to precision health research. It is critical to understand target communities' motivations and barriers to participation in research.

Delivering culturally appropriate text messages via age-appropriate texting and social media platforms may be an effective recruitment and intervention strategy. The next step is to develop and examine the feasibility of a culturally tailored precision health texting strategy for Vietnamese Americans.

(JMIR Mhealth Uhealth 2021;9(3):e23058) doi:10.2196/23058

KEYWORDS

Vietnamese Americans; texting; precision health; qualitative research; mobile phone

Introduction

Background

Precision health is defined as *an innovative approach that considers individual differences in people's genes, environments, and lifestyles* [1]. The largest effort in precision health research is the Precision Medicine Initiative (PMI or the *All of Us Research Program),* which aims to include 1 million or more participants to be a part of a diverse database that can help revolutionize precision health research [2,3]. The official launch of the *All of Us Research Program* was spring 2018 [4] and is anticipated to last for at least a decade. According to the All of Us Research Hub, which provides aggregate public data about the participants, the PMI currently includes 225,140 participants, as of February 11, 2020 [5].

However, when visiting the *All of Us* website (as of May 3, 2020) [6], it does not appear that non–English-speaking and non–Spanish-speaking racial and ethnic minorities, particularly Vietnamese Americans and persons with limited English proficiency (LEP), have been engaged. For example, there were no translated materials and videos in Vietnamese on the website. Moreover, when visiting the website [6], it appears that to enroll in the program, one needs to have an email address (and therefore access to the internet and a device such as a computer or smartphone).

It is important to ensure that underserved communities, defined by the Department of Health and Human Services as "communities that include members of minority populations or individuals who have experienced health disparities" [7], are engaged and educated on precision health. Asian Americans and Pacific Islanders (AAPI), for example, experience significant health and health care disparities [8], especially given that AAPI comprises 6% of the US population and is the fastest growing racial group in the United States [9]. They are heterogeneous in terms of languages and dialects (>100), cultural groups (>50), immigration patterns, religions, diets, and socioeconomic status [10]. Specifically, Vietnamese Americans have encountered premigration and migration traumas and other barriers (eg, low socioeconomic position, language difficulties) that place them at a higher risk for physical and mental illnesses compared with the general population [11-13]. As a relatively recent immigrant group, Vietnamese Americans are likely to be either first-generation or second-generation Americans, and there are about 2 million Vietnamese Americans in the United States [14]. They are the fourth largest foreign-born population from Asia (after China, India, and the Philippines) and have high rates of naturalization: about 76% of foreign-born Vietnamese Americans are naturalized US citizens.

Despite the growing AAPI population and the ongoing PMI, research participation from AAPI including Vietnamese Americans is limited. Of the 225,140 All of Us participants (database version February 11, 2020; 740/225,140, 0.33%) identified as Vietnamese [5]. Existing, limited research shows that AAPI are interested in being engaged but that various community concerns, such as a lack of culturally and linguistically appropriate information, need to be addressed to improve participation in clinical research and that engagement messages should include an *offer of hope* [15]. Moreover, a recent paper outlines the opportunities and challenges in precision medicine in improving cancer prevention and treatment for Asian Americans [16]. An integral point of their paper is to ensure that personalized medicine becomes a reality for all Americans [16].

Objectives

Given the goals of the PMI to recruit 1 million persons into a precision health study of ≥10 years and that such persons should reflect the cultural diversity in the United States, it is important to explore strategies to meaningfully engage and educate racial and ethnic minorities about precision health and precision health research. Specific factors, such as English language proficiency, warrant special attention. Limited research exists on the feasibility or acceptability of using widely available mobile tools, such as texting for persons with LEP, including AAPI, for PMI-related purposes. Although traditional recruitment methods such as ethnic-specific radio/television advertisements have been successful in reaching out to Vietnamese Americans [17,18], this method has its limitations. The limitations include (1) not reaching every Vietnamese American as these audiences tend to be older in demographics, thereby limiting the ability to reach the younger generations; (2) high costs; and (3) the inability to reach the intended audience in *real time*, as it requires significant coordination with the radio/television producers/stations to get the messages out. It is critical to use a range of outreach methods, including those that are more cost-effective and wide reaching, to engage and educate Vietnamese Americans about precision health and precision health research. Accordingly, this study aims to examine the attitudes and perceptions of precision health, motivations and barriers to participate in precision health research, and acceptability of texting as a recruitment and intervention strategy among underserved English-speaking and Vietnamese-speaking Vietnamese Americans.

Methods

Recruitment

A community sample of 37 Vietnamese Americans comprising 3 age groups (18-30, 31-59, and ≥60 years) was recruited and participated in the study. During a 2-week period, 3 recruitment strategies were used: (1) announcements were made to the local community organizations that serve Vietnamese Americans, (2) flyers and word of mouth, and (3) Vietnamese language radio announcements. Given that one might have heard from multiple sources, 40 people reported that they heard about the study from friends or family (ie, word of mouth) or saw a flyer. A total of 30 people mentioned that they received announcements or heard directly through the primary Vietnamese community–based organization (ie, International Children Assistance Network [ICAN]), and 10 people learned about the study recruitment through a Vietnamese radio channel. The remaining channels were also reported: Facebook (n=3), email (n=3), and other (from a community outreach event; n=1).

A total of 73 persons expressed their interest to participate in the study. Out of the 73 people, 3 were not eligible for the following reasons: not available on the date of their age group (n=1), aged below 18 years (n=1), and called in to register after their age group had taken place (n=1). Of the 70 who were screened eligible to participate in the study, 33 did not participate because of the following reasons: withdrew owing to a last-minute schedule conflict (n=8), felt uncomfortable talking and answering questions in a focus group (3), had no transportation to the focus group (n=1), or wait-listed because the study reached its maximum number of preferred participants for the focus groups (n=21). The remaining 37 eligible participants were included in the study.

Quantitative Data

After obtaining informed consent, participants were asked to complete a questionnaire pertaining to their sociodemographic background; questions were related to their knowledge, attitudes and behaviors in genetics and genetic testing, and texting attitudes for precision health. In total, 6 participants were able to fill their surveys directly on a web-based survey tool (ie, REDCap survey); however, the remaining 31 completed their surveys on hard copies because of a lack of internet access and/or a limited understanding on how to complete a web-based survey. Research staff later entered hard copy data onto REDCap.

Sociodemographic and Texting Experience Questionnaire

We asked participants about their sociodemographic background, including their race, ethnicity, gender, year of birth, nativity, years lived in the United States, marital status, type of medical insurance, employment, education, household income and size, and English language proficiency.

Texting experience comprised questions, such as whether and how often they used a phone to send text messages, whether they used a phone to receive text messages, and which texting apps they used on their phones. We also asked participants if they had prior research participation experience.

Genetics and Genetic Testing Survey

Participants were asked to indicate to what extent they agreed with the statements in the survey on a 5-point Likert scale from *strongly disagree* (*1*) to *strongly agree* (*5*). These items, which were drawn from previous precision health research with diverse racial/ethnic populations including Vietnamese Americans [18,19], included (1) genetics affect health; (2) the use of personal genetic information in health care is beneficial to patients; (3) I feel comfortable talking to my doctors about genetic testing; (4) the use of genetic tests could lead to discrimination; (5) if available, I would like to undergo genetic testing to find out if I am at risk of certain diseases (eg, cancer and diabetes); (6) if available, I would like to undergo genetic testing to find out if a certain drug (eg, high blood pressure medication) or treatment would work for my conditions; and (7) if available, I would like to participate in research about using genetics for disease prevention and treatment.

Attitudes About Using Texting for the Precision Health Education Survey

Participants were asked to indicate to what extent they agreed with 3 statements to assess their attitudes about using texting for precision health education on a 5-point Likert scale from *strongly disagree* (*1*) to *strongly agree* (*5*). These items, which were developed for this study, included (1) I believe that text messages can help the community to understand how genetics affects health; (2) I believe that text messages can help the community to get genetic testing; and (3) I believe that text messages can help the community to understand what precision health is.

Qualitative Data: Focus Groups

Three focus groups, which were separated by age groups, were conducted by the same bilingual research staff (QV and KN) in either English or Vietnamese. The focus group for participants aged 18 to 30 years (youngest age group) was conducted in the English language with 12 participants. The 2 focus groups for participants aged 31 to 59 years (middle-age group) and ≥60 years (oldest age group) were conducted in Vietnamese with 12 and 13 participants, respectively.

The facilitator (QV) from ICAN has significant prior experience facilitating focus groups. She began the focus group discussion by asking participants to share their knowledge regarding precision health and precision health research. After gauging their understanding, the facilitator explained what precision health and precision health research are. She then guided the focus group discussion. Each focus group lasted about 90 min, and the participants received a US $25 gift card to a local store for their time.

During the focus group, participants were asked to share (1) their knowledge of precision health and genetics; (2) knowledge, attitudes, and experiences with research; (3) perspectives and experiences with texting in general and with various texting apps; (4) insights on how to frame text messages on precision health for Vietnamese Americans; and (5) recommendations to best outreach to the Vietnamese community regarding precision health.

Translation Process and Validity

The following materials were translated into Vietnamese: recruitment flyers, screening eligibility questions, informed consent, sociodemographic questionnaire, texting survey items, and focus group guide. The World Health Organization's guidelines on translation and adaptation of instruments [20] were used to guide the translations of the study materials. By using this established translation method such as forward and backward translation methods, we were able to attain *conceptually equivalent* Vietnamese language versions of the English language materials, which were translated to be *cross-cultural and conceptual, rather than to be of linguistic/literal equivalence* [20]. Bilingual and bicultural Vietnamese research staff conducted the translations.

Quantitative Data Analysis

Descriptive statistics were provided for the overall sample and focus group participation (age groups). Chi square and analysis of variance statistics were used to determine statistically significant differences among the groups. The data were analyzed using STATA 16 software [21].

Qualitative Data Analysis

The focus group recordings were translated and transcribed verbatim into Microsoft Word documents. To increase the accuracy of the qualitative data, the research assistants conducted word-for-word translations of the focus group from Vietnamese to English. Content analysis was conducted on these transcripts by 2 raters. Using the questionnaire and common topics covered by the facilitator during each focus group session, a coding dictionary was created to be used as a guide. Subsequently, the 2 raters independently conducted thematic coding of the qualitative data using the thematic analysis approach by Luborsky [22]. After the initial content analysis, the raters discussed discrepancies in coding and agreed upon the emergent themes and subthemes (Textbox 1). The summary and reporting of the qualitative data were guided by examples from the literature on how to present qualitative findings [23] and examples from previous research that used qualitative methods to study vulnerable populations [24,25].

Human Subjects Protection

This research was approved by the Stanford University Institutional Review Board (protocol no. 51409). Informed consent was obtained from all participants before study participation.

Results

Sample Characteristics

A total of 37 Vietnamese Americans participated in 3 focus groups by age groups: 18 to 30, 31 to 59, and ≥60 years (Table 1). Overall, slightly more females than males participated, but there was no significant difference by age group for gender. Overall, more than half (21/37, 57%) were married or living with a partner, with the remaining reporting being single (11/37, 30%) or separated/divorced/widowed (5/37,14%). There were significant differences by marital status for the age groups, with a high proportion (88%) of the youngest age group (18-30 years) reporting being single.

Approximately three-quarters (29/37, 78%) of the participants had acquired at least some college education. Regarding employment status, more than half (21/37, 57%) reported that they were employed either full time or part time, with the remaining stating that they were disabled/retired (10/37, 27%), a homemaker (3/37, 8%), unemployed (2/37, 5%), or a student (1/37, 3%). Household income was also reported, with more than half (22/37, 59%) of the participants making less than US $25,000 in 2018 for the entire household, and this stayed with the highest proportions (above 50%) among all 3 age groups.

There were significant differences by nativity, with most participants being born in Vietnam (34/37, 92%). Three US-born participants belonged to the youngest age group. About half (20/37, 54%) reported that they could speak at least some English and about half (17/37, 46%) could speak fluent or native English. The youngest age group tended to be more fluent in the English language and was the only group with participants who self-identified as native English speakers. The level of English fluency was affected by age groups. For instance, a high proportion (10/12, 83%) of the youngest age group stated being fluent in the English language, whereas only 33% (4/12) of the middle-aged participants and 25% (3/12) of the older age group stated this. On the contrary, when reporting English fluency as *some English*, the older age group (10/13, 77%) and middle-aged group (8/12, 67%) tended to have higher proportions.

Most participants (27/37, 73%) reported that they sent text messages regularly, with the highest proportion (12/12, 100%) from the youngest age group who reported that their phones can receive text messages (36/37, 97%). Phone texts (n=26) and Facebook (n=31) were the most common texting apps used among the participants, other apps included Viber (n=18), WhatsApp (n=5), WeChat (n=3), and others (n=8).

Table 1. Sample characteristics by age groups (N=37).

Characteristics	Total (N=37)	Age group (years)			*P* value
		18-30 (n=12)	31-59 (n=12)	≥60 (n=13)	
Age (years)					<.001
Mean (SD)	48.8 (19.3)	25.8 (3.0)	49.3 (10.1)	69.5 (4.4)	
Median	52	27	53.5	68	
Range	21-76	21-29	32-59	62-76	
Gender, n (%)					.62
Female	22 (59.5)	6 (50)	7 (58.3)	9 (69.2)	
Male	15 (40.5)	6 (50)	5 (41.7)	4 (30.8)	
Current marital status, n (%)					<.001
Single	11 (29.7)	10 (83.3)	1 (8.3)	0 (0)	
Married/living together	21 (56.8)	2 (16.7)	9 (75)	10 (76.9)	
Separated/divorced/widowed	5 (13.5)	0 (0)	2 (16.7)	3 (23.1)	
Education, n (%)					.15
Less than high school	2 (5.4)	0 (0)	2 (16.7)	0 (0)	
High school	6 (16.2)	2 (16.7)	3 (25)	1 (7.7)	
Some college	10 (27)	2 (16.7)	1 (8.3)	7 (53.8)	
College or more	15 (40.5)	5 (41.7)	5 (41.7)	5 (38.5)	
Graduate	4 (10.8)	3 (25)	1 (8.3)	0 (0)	
Employment status, n (%)					<.001
Full time	15 (40.5)	9 (75)	6 (50)	0 (0)	
Part time	6 (16.2)	1 (8.3)	3 (25)	2 (15.4)	
Homemaker	3 (8.1)	0 (0)	2 (16.7)	1 (7.7)	
Unemployed	2 (5.4)	1 (8.3)	1 (8.3)	0 (0)	
Disabled/retired	10 (27)	0 (0)	0 (0)	10 (76.9)	
Student	1 (2.7)	1 (8.3)	0 (0)	0 (0)	
Household income (US $), n (%)					.65
<25,000	22 (59.5)	6 (50)	7 (58.3)	9 (69.2)	
25,000-75,000	9 (24.3)	3 (25)	3 (25)	3 (23.1)	
75,001-150,000	5 (13.5)	3 (25)	1 (8.3)	1 (7.7)	
>150,000	1 (2.7)	0 (0)	1 (8.3)	0 (0)	
Nativity, n (%)					.03
US-born	3 (8.1)	3 (25)	0 (0)	0 (0)	
Foreign-born	34 (91.9)	9 (75)	12 (100)	13 (100)	
English fluency, n (%)					.006
Native English	4 (10.8)	4 (33.3)	0 (0)	0 (0)	
Fluent English	13 (35.1)	6 (50)	4 (33.3)	3 (23.1)	
Some English	20 (54.1)	2 (16.7)	8 (66.7)	10 (76.9)	
Send text messages, n (%)					.12
Regularly	27 (73)	12 (100)	8 (66.7)	7 (53.8)	
Sometimes	8 (21.6)	0 (0)	3 (25)	5 (38.5)	
Not often	2 (5.4)	0 (0)	1 (8.3)	1 (7.7)	
Receive text messages, n (%)					.39

Characteristics	Total (N=37)	Age group (years)			*P* value
		18-30 (n=12)	31-59 (n=12)	≥60 (n=13)	
Yes	36 (97.3)	12 (100)	12 (100)	12 (92.3)	
No	1 (2.7)	0 (0)	0 (0)	1 (7.7)	
Types of text messages[a], n (%)					.09
Phone texts	26 (70.3)	11 (91.7)	9 (75)	6 (46.2)	
WeChat	3 (8.1)	2 (16.7)	1 (8.3)	0 (0)	
WhatsApp	5 (13.5)	3 (25)	2 (16.6)	0 (0)	
Facebook	31 (83.8)	11 (91.7)	12 (100)	8 (61.5)	
Viber	18 (48.6)	4 (33.3)	11 (91.7)	3 (23.1)	
Other[b]	10 (27)	5 (41.7)	2 (16.6)	3 (23.1)	

[a]Multiple-answer question.

[b]Other: Zalo, Snapchat, Skype, Line, Instagram, and Discord (n=10).

Overall and by Genetic Testing and Texting

Table 2 shows the knowledge-attitude-behavior (KAB) scores within each group. The mean KAB score ranged from 1 (strongly disagree) to 5 (strongly agree). The overall mean KAB score was 3.95 (SD 0.36). The mean genetic testing score was 3.94 (SD 0.41), and the mean texting summary score was 3.96 (SD 0.63). There were no significant differences by age groups.

Table 2. Mean difference in summary scores by age groups: overall, genetics and genetic testing, and texting (N=37).

Category	Total (N=37)	Age group (years)			*P* value
		18-30 (n=12)	31-59 (n=12)	≥60 (n=13)	
Overall, mean (SD)	3.95 (0.36)	4.00 (0.98)	4.00 (0.10)	3.86 (0.11)	.55
Genetics, mean (SD)	3.94 (0.41)	3.96 (0.13)	3.96 (0.09)	3.91 (0.13)	.35
Texting, mean (SD)	3.96 (0.63)	4.08 (0.18)	4.08 (0.18)	3.74 (0.61)	.09

Qualitative Findings

The main themes from the qualitative data included the following: (1) perceptions of precision health, (2) motivation to participate in the research study, (3) advantages of texting, (4) barriers to texting, and (5) recommendations for texting strategies for precision health with Vietnamese Americans (Textbox 1).

Textbox 1. Overall qualitative themes and subthemes.

Perceptions of precision health

- Genetic predisposition
- Lack of knowledge on precision health
- Usage as preventative measure
- Environmental factors

Motivation to participate in research study

- Representation of the Vietnamese American population
- Desire for knowledge
- Desire to contribute to research
- Previous experiences in participating in research studies
- Incentives
- Credibility of the research study or researcher

Advantages of texting

- Time efficiency and convenience
- Group messages
- Ease of sharing multimedia

Barriers to texting

- Misinterpretation
- Cost and/or need for Wi-Fi or mobile data
- Differing texting platforms among age groups
- Spam messages

Recommendations for texting strategies

- Short and concise messages
- Frequency—at most 1 to 2 messages per week
- Inclusion of pictures or links with brief descriptions
- Maintain respondents' confidentiality
- Flexibility in language selection
- Simple language; use minimal medical, professional terminology (in English if necessary)
- Ability to forward or share information from texts
- Credible organizations and sources

Perceptions of Precision Health

Qualitative findings revealed that several participants expressed their perceptions of precision health and stated that precision health involves the study of genetic predisposition and use of genes to find specific treatments. The perceptions of precision health were similar in all 3 age groups, as participants discussed the association between family history, genes, and diseases.

Although some participants had previously heard of precision health, other participants reported a lack of knowledge in precision health and desired to learn about the meaning of precision health. Lack of knowledge on precision health was found in all 3 age groups; however, it was more prevalent in the youngest age group. Participants discussed that precision health can be used as a tool to prevent certain diseases. The discussions in all 3 age groups revealed participants' understanding of precision health as a preventative measure to reduce the risk of diseases. In addition, some (n=5) reported that environmental factors, such as occupation and lifestyle, play important roles in determining a person's health.

Motivation to Participate in Research

The participants reported that they were motivated to participate in research studies because they wanted more Vietnamese American representation in the field of research. One participant stated:

> *...but the most important thing is this research study is for Vietnamese Americans and Asian Americans. We (need to) participate to have a voice in the area of research so that researchers can learn what diseases are relevant to Vietnamese Americans and Asian Americans so they can tailor the treatments appropriately.* [Participant 17, female, 52 years old]

The participants believed that more representation of Vietnamese Americans was needed. The motivation to participate in research studies also reflected the participants' desire for more knowledge in the research topics (n=13). The desire for knowledge was relevant in all 3 age groups. Previous participation in research studies also motivated several participants (n=4) to participate in this particular research study. Some participants (n=5) stated that incentives also influenced their motivation to participate in this study. Finally, a few (n=2) participants stated that the credibility of the research (ie, trusting the researcher, community-based organization, and/or university) may motivate them to participate in the research.

Advantages of Texting

The most common advantage noted across all 3 focus groups was the convenience and time efficiency of texting (*n*=12). One participant summarized the effectiveness of texting:

> *I would say we get an instant alert right away. If something happens, it's just one text that we can get the information right away. I prefer texting over calling nowadays because I'm always busy with doing something, so I'd rather have someone text me and I'd get back to them right away.* [Participant 10, male, 27 years old]

Two participants noted their preferences for texting over sending emails:

> *I only check emails every 2-3 days. For texts, when we hear ding, we (can) open and read immediately.* [Participant 37, female, 67 years old]

> *I feel like when I get a text, it's for me, so I prefer texts than emails.* [Participant 31, female, 64 years old]

Other common advantages included features of text messages, such as group messaging (n=2), confidentiality (n=1), and the ease of sharing multimedia via texting (*n*=3). Interestingly, a participant from the age group 18 to 30 years noted the sense of autonomy that texting provided:

> *Just going back to it fits your schedule. If you get a text and you want to reply to it, you can. If you don't want to reply to it, you can put it off and lie, say you didn't get it. Or just not respond.* [Participant 8, male, 22 years old]

From the older age group (≥60 years) in particular, a few participants (n=3) remarked on the more widespread access to cellphones than other devices such as computers.

Barriers to Texting

Several participants (n=5) reported the misinterpretation of text messages by the receiver as a barrier. Causes of misinterpretation included being unable to convey the tone of voice that would normally be conveyed if having a conversation in person. Text messages may also be disregarded as spam messages. Participants (n=3) from the older age group also brought up the varying use of tone symbols when typing in Vietnamese and abbreviated words as causes for misinterpretation. One participant provided an example stating:

> *Especially Vietnamese, people would abbreviate "m i ng i" as "mn", I don't know if they mean "mi n nam" ("South") or something else.* [Participant 37, female, 67 years old]

In addition, participants (n=5) noted that the use of varying texting platforms could create difficulty in texting people, even among just the Vietnamese American population. One of the participants stated:

> *I would say it depends on the group of people you try to reach out to as I know my parents would use Viber, but they don't text, for some reason. But it's pretty much the same method of sending messages out. They prefer to use Viber because their friends are using Viber. For me I don't really text with the (mobile) carrier, but I use Facebook messenger. So, it really depends on the group that you hang out with or that you work with. I'd prefer a different way of sending out messages.* [Participant 10, male, 27 years old]

When asked which texting services were used by the participants in each focus group, a list of apps was compiled, including Facebook Messenger, Viber, Zalo, Line, Google Voice, WhatsApp, Kakao Talk, and mobile carrier texts.

Concerns about cost and the need for Wi-Fi and/or mobile data were also suggested as barriers to texting (n=5). One participant from the youngest age group stated:

> *I'm actually annoyed by those internet-based apps because you always have to have Wi-Fi or your data on and if you don't then there is a serious delay to the message. So, the contacts where we both have messenger, I don't get their messages or otherwise unless the data is on or I have access to internet.* [Participant 9, male, 28 years old]

Other barriers put forth by the participants were having to switch between English and Vietnamese keyboards (n=1), autocorrect (n=2), and the lack of knowledge regarding how to set up different texting apps (n=1).

Recommendations for Texting Strategies

Several participants (n=8) from all 3 age groups discussed the need for a concise, simple language when texting health information. They recommended the use of minimal medical or professional terminology for one of the texting strategies. One participant stated:

> *I'd say use simple language because he mentioned earlier that some medical terminology might be too complicated for someone to understand what it means. So, try to use a daily language, it'd be helpful.* [Participant 10, male, 27 years old]

> *Don't use too many medical terms.* [Participant 21, male, 32 years old]

In addition, participants (n=8) stated that the frequency of text reminders should not exceed more than 1 or 2 text messages in 1 week. If the text messages were sent frequently, the participants were likely to consider the text messages irrelevant. One participant from the youngest age group stated:

> *I think it's also very important who the message is coming from...But if it was another entity, or I had no idea who they were, I would have felt annoyed at the frequency because I got 3 messages as a reminder for this and an email.* [Participant 7, female, 28 years old]

Participants (n=2) from the older age group emphasized the inclusion of pictures, links, and brief descriptions of relevant health information when sending text messages.

Several participants (n=6) expressed concerns regarding confidentiality and security when receiving text messages. They stated that it is necessary to maintain confidentiality when sending text messages involving personal information. This concern was common in all 3 age groups, as 1 participant from the youngest age group reported:

> *Like if you send the important information, health conditions of one person through texting that is kinda not formal. I don't feel it's safe to send that information through texts.* [Participant 6, female, 26 years old]

Another recommendation is language selection. Several participants (n=6) suggested the need to offer language preferences because of the diversity in experience and background. However, when using Vietnamese, it needed to be written correctly in grammar and form to avoid misunderstanding. The facilitator confirmed this understanding when she stated, "The content should be in Vietnamese, but the terminology would be also included in English." One participant replied, "That's correct" (participant 31, female, 65 years old). Another participant added, "Vietnamese with tones" (participant 30, male, 73 years old).

In addition, when the focus group facilitator asked the participants (in the 60 years old and older focus group) on whether medical terminology should be translated, they responded, "No need to translate." In other words, it is best to keep its given name and translate the concept instead. This was further supported by participants in the other focus groups (18-30 and 31-59 years old):

> *If the translation seems odd or tricky, then include the English term.* [Participant 15, female, 58 years old]

> *There are a lot of direct translations of English like legal or judicial or medical terms sometimes they are very off.* [Participant 9, male, 28 years old]

Participants (n=2) from the middle-aged group discussed their ability to forward or share relevant information from texts as one of the texting strategies. One participant stated:

> *When I see health information on Facebook, I would select the ones that would be of great benefit to me and then I would share them and keep them for myself...If I see them as helpful, such as exercise for elders or new treatments. Of course I always select them carefully if I want to keep for myself and forward to friends.* [Participant 18, female, 59 years old]

Discussion

Summary of Findings and Comparison With Previous Research

This pilot study is the first of its kind to examine the attitudes and perceptions of precision health and precision health research while also exploring a novel method of recruitment (ie, texting) among an underserved racial population in the United States, especially Vietnamese Americans. Specifically, the study explored motivations and barriers to research participation and cultural acceptability of texting as a recruitment and intervention strategy for precision health research.

The overall findings show that although many of the participants reported being familiar with genetics and genetics testing, some reported lack of knowledge in precision health or precision medicine and desired to learn more about these topics. Once the participants were more informed about precision health (through the facilitator's introduction of the topics), they became more receptive and expressed positive attitudes toward participating in precision health research and provided input on how texting could be used to recruit and engage Vietnamese Americans in research. Although there is limited prior research on this topic, there was 1 qualitative study (where the lead author of this paper was a collaborator) that had also employed focus groups among American Indian, African American, Latino, Chinese, and Vietnamese groups and found similar findings (ie, limited prior knowledge about precision health) [19]. Moreover, the same study reported that all Latino, Chinese, Vietnamese, and some African American participants had positive attitudes toward precision health [19].

Qualitative themes indicated that although it slightly varied among the 3 age groups, most participants were motivated to participate in research because of the lack of the Vietnamese American representation in the area of research, and they expressed how they would want to contribute to this effort (to increase research participation). In addition, many participants reported a desire to gain more knowledge in their research.

The study findings support the notion that to engage Vietnamese Americans in research (eg, precision health research), the community's needs must be met. For example, Vietnamese Americans need to understand the specific benefits of research participation personally and for their communities [26]. Moreover, research needs to be culturally tailored for Vietnamese Americans in language and cultural beliefs/norms [26,27].

Participants responded positively to using texting as a recruitment strategy for precision health research. They indicated that they prefer texting over other communication methods (eg, emails) because it is convenient, time efficient, low cost, and widely used. About three-quarters of the study participants reported that they sent text messages regularly, and most of them received text messages. The 2 most commonly reported

texting methods were Facebook (n=31) and phone texts (n=26). Although previous research has not focused on texting as a strategy for precision health research, it has highlighted the acceptability of text messages for delivering health messages to different ethnic minority populations [28-30]. Furthermore, a recent systematic review found that few studies have examined the use of digital tools (eg, texting) for the recruitment and retention of racial/ethnic minority participants in intervention research and that randomized controlled trials are limited. Thus, the next phase of our research may play an important role in contributing to our understanding of the potential utility of texting as a precision health research recruitment tool for racial/ethnic minorities such as Vietnamese Americans.

The participants offered recommendations to overcome potential barriers to implement texting as a recruitment strategy for precision health research. For example, a common barrier that participants brought up was the misinterpretation of the meanings or intentions of the text messages, including a lack of tone of voice and the use of abbreviations, which are important in the Vietnamese language [31]. Moreover, it is important to select words that are easy to understand, and if using Vietnamese, words need to be culturally appropriate and have tone symbols, which is particularly relevant when providing appropriate services to people who speak Vietnamese[31]. These recommendations are consistent with previous research on other racial/ethnic populations. For example, the use of abbreviations, particularly abbreviated slang, would be difficult to understand among older Samoans [30].

Strengths and Limitations

A strength of our study is that our focus groups were separated by age groups (18-30, 31-59, and 60 years and older) and language proficiency (English; Vietnamese). This enabled us to ensure diverse perspectives by age (young adults, middle-age adults, and older adults) and Vietnamese proficiency with LEP. Future research could involve focus groups being conducted with the same gender (ie, all males in one focus group and females in a different focus group) and/or semistructured interviews to potentially elicit shared perspectives by gender and/or permit the participants to provide detailed perspectives.

This study had several limitations. First, we conducted a qualitative study on a small sample of Vietnamese Americans. The data presented might not be generalizable to other Vietnamese Americans in the United States. Second, the discussion on using mobile tools for recruitment and research was focused more on texting and social media platforms, although other strategies might also be viable. Third, the feasibility and acceptability of using texting as a recruitment and intervention strategy were based on self-report attitudes in focus groups, which merits field testing to confirm the feasibility of conducting precision health research in the underserved Vietnamese American community.

Conclusions

To improve the research participation among racial/ethnic minority populations and have equitable representation that mirrors the diversity of the United States in precision health research, there needs to be more thoughtful and committed efforts to engage them. These efforts include having culturally and linguistically tailored materials, age-appropriate texting strategies, and meaningful benefits to the Vietnamese American community. Furthermore, it is important to explore alternate ways of engaging with potential participants who do not have access to digital technology, such as email. Precision health research, and all research more broadly, which requires an email address for participation, would exclude many potential participants in marginalized populations. Our study provides support for the use of texting as an example of an alternate strategy to engage such participants.

As a next step of this formative research, we will be developing and piloting a culturally tailored texting strategy as an educational and engagement tool for Vietnamese Americans about precision health. In the next study phase, we will examine the acceptability of such a texting strategy and develop a library of culturally appropriate precision health messages that may be implemented on a large scale. We will partner with a large Vietnamese community–based organization for this important next phase of community-based participatory research.

Acknowledgments

This research was supported by the National Institute on Minority Health and Health Disparities of the National Institutes of Health (NIH) under Award Number U54MD010724. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The authors would like to thank the participants for their participation. Dr Periyakoil's work was supported by NIH grants P30 AG059307, U54 MD010724, and R01 AG062239.

Authors' Contributions

Conceptualization was carried out by VTP; methodology by VMTP; formal analysis by AK, IC, and BN; investigation by VTP; original draft preparation (writing) by VTP, AK, IC, BN, KN, and YAH; review and editing (writing) by VTP, AK, IC, BN, KN, QV, YH, and VP; supervision by VTP; project administration by VMTP, QV, and KN; and funding acquisition by VTP, QV, and VP.

Conflicts of Interest

None declared.

References

1. The Precision Medicine Initiative. Obama White House. 2015. URL: https://obamawhitehouse.archives.gov/precision-medicine [accessed 2017-03-05]
2. NIH awards $55 million to build million-person precision medicine study. National Institutes of Health. 2016 Jul 06. URL: https://www.nih.gov/news-events/news-releases/nih-awards-55-million-build-million-person-precision-medicine-study [accessed 2018-03-21]
3. Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med 2015 Feb 26;372(9):793-795. [doi: 10.1056/NEJMp1500523] [Medline: 25635347]
4. All of Us Research Program FAQs. National Institutes of Health. 2018. URL: https://allofus.nih.gov/about/faq [accessed 2018-03-21]
5. All of Us Research Hub. National Institutes of Health. 2020. URL: https://databrowser.researchallofus.org/survey/the-basics [accessed 2020-11-11]
6. National IOH. All of Us Research Program. National Institutes of Health. 2020. URL: https://allofus.nih.gov/ [accessed 2020-05-03]
7. Serving Vulnerable and Underserved Populations. Centers for Medicare & Medicaid Services. 2020. URL: https://marketplace.cms.gov/technical-assistance-resources/training-materials/vulnerable-and-underserved-populations.pdf [accessed 2020-11-13]
8. Health Disparities. Asian Health Coalition. URL: https://www.asianhealth.org/our-voices/health-disparities/ [accessed 2018-03-21]
9. Facts for Features: Asian/Pacific American Heritage Month: May 2018. U.S. Census Bureau. 2018 May 01. URL: https://www.census.gov/newsroom/facts-for-features/2018/asian-american.html [accessed 2018-08-28]
10. Facts for Features: Asian/Pacific American Heritage Month: May 2011. U.S. Census Bureau. 2011 Apr 29. URL: https://www.census.gov/newsroom/releases/archives/facts_for_features_special_editions/cb11-ff06.html [accessed 2011-09-09]
11. Hinton WL, Tiet Q, Tran CG, Chesney M. Predictors of depression among refugees from Vietnam: a longitudinal study of new arrivals. J Nerv Ment Dis 1997 Jan;185(1):39-45. [doi: 10.1097/00005053-199701000-00007] [Medline: 9040532]
12. Who are Asian Americans? Asian American Health Initiative. URL: https://www.aahiinfo.org/english/asianAmericans.php [accessed 2018-03-21]
13. Ta Park VM, Goyal D, Nguyen T, Lien H, Rosidi D. Postpartum traditions, mental health, and help-seeking considerations among vietnamese american women: a mixed-methods pilot study. J Behav Health Serv Res 2017 Jul;44(3):428-441. [doi: 10.1007/s11414-015-9476-5] [Medline: 26276422]
14. The Vietnamese Population in the United States: 2010. U.S. Census Bureau. 2011. URL: http://www.vasummit2011.org/docs/research/The%20Vietnamese%20Population%202010_July%202.2011.pdf [accessed 2014-02-25]
15. Tzuang M, Gallagher-Thompson D. Caring for Caregivers of a Person with Dementia. In: Pachana NA, Laidlaw K, editors. The Oxford Handbook of Clinical Geropsychology: International Perspectives. Oxford, UK: Oxford University Press; 2014.
16. Trinh-Shevrin C, Sacks R, Ahn J, Yi SS. Opportunities and challenges in precision medicine: improving cancer prevention and treatment for Asian Americans. J Racial Ethn Health Disparities 2018 Feb;5(1):1-6 [FREE Full text] [doi: 10.1007/s40615-016-0334-9] [Medline: 28127673]
17. Ta Park VM, Ton V, Tiet QQ, Vuong Q, Yeo G, Gallagher-Thompson D. Promising results from a pilot study to reduce distress in Vietnamese American dementia and memory loss caregivers. Alzheimers Dement (N Y) 2019;5:319-327 [FREE Full text] [doi: 10.1016/j.trci.2019.05.006] [Medline: 31384663]
18. Ta Park VM, Olaisen RH, Vuong Q, Rosas LG, Cho MK. Using Korean dramas as a precision mental health education tool for Asian Americans: a pilot study. Int J Environ Res Public Health 2019 Jun 18;16(12):- [FREE Full text] [doi: 10.3390/ijerph16122151] [Medline: 31216626]
19. Rosas LG, Nasrallah C, Park VT, Vasquez JJ, Duron Y, Garrick O, et al. Perspectives on precision health among racial/ethnic minority communities and the physicians that serve them. Ethn Dis 2020;30(Suppl 1):137-148 [FREE Full text] [doi: 10.18865/ed.30.S1.137] [Medline: 32269455]
20. Process of translation and adaptation of instruments. World Health Organization. 2018. URL: https://www.who.int/substance_abuse/research_tools/translation/en/ [accessed 2018-09-11]
21. Stata Statistical Software: Release 16. StataCorp. College Station, Texas, USA: Stata Corporation; 2019. URL: https://www.stata.com/ [accessed 2020-11-14]
22. Luborsky M. The identification and analysis of themes and patterns. In: Gubrium JF, Sankar A, editors. Qualitative methods in aging research. Thousand Oaks, California: SAGE Publications, Inc; 1994:189-210.
23. Anderson C. Presenting and evaluating qualitative research. Am J Pharm Educ 2010 Oct 11;74(8):141 [FREE Full text] [doi: 10.5688/aj7408141] [Medline: 21179252]
24. Goyal D, Park VT, McNiesh S. Postpartum depression among Asian Indian mothers. MCN Am J Matern Child Nurs 2015;40(4):256-261. [doi: 10.1097/NMC.0000000000000146] [Medline: 26121757]
25. Kim K, Kim B, Choi E, Song Y, Han H. Knowledge, perceptions, and decision making about human papillomavirus vaccination among Korean American women: a focus group study. Womens Health Issues 2015;25(2):112-119 [FREE Full text] [doi: 10.1016/j.whi.2014.11.005] [Medline: 25747518]

26. Ta Park V, Nguyen K, Tran Y, Yeo G, Tiet Q, Suen J, et al. Perspectives and Insights from Vietnamese American Mental Health Professionals on How to Culturally Tailor a Vietnamese Dementia Caregiving Program. Clin Gerontol 2018;41(3):184-199. [doi: 10.1080/07317115.2018.1432734] [Medline: 29419364]
27. Ta Park VM, Ton V, Yeo G, Tiet QQ, Vuong Q, Gallagher-Thompson D. Vietnamese American dementia caregivers' perceptions and experiences of a culturally tailored, evidence-based program to reduce stress and depression. J Gerontol Nurs 2019 Sep 01;45(9):39-50. [doi: 10.3928/00989134-20190813-05] [Medline: 31437289]
28. Prinjha S, Ricci-Cabello I, Newhouse N, Farmer A. British south Asian patients' perspectives on the relevance and acceptability of mobile health text messaging to support medication adherence for type 2 diabetes: qualitative study. JMIR Mhealth Uhealth 2020 Apr 20;8(4):e15789 [FREE Full text] [doi: 10.2196/15789] [Medline: 32310150]
29. Dobson R, Whittaker R, Bartley H, Connor A, Chen R, Ross M, et al. Development of a culturally tailored text message maternal health program: textmatch. JMIR Mhealth Uhealth 2017 Apr 20;5(4):e49 [FREE Full text] [doi: 10.2196/mhealth.7205] [Medline: 28428159]
30. McCool J, Tanielu H, Umali E, Whittaker R. Assessing the cross-cultural adaptation and translation of a text-based mobile smoking cessation program in Samoa (txttaofitapaa): pilot study. JMIR Mhealth Uhealth 2018 Aug 31;6(8):e173 [FREE Full text] [doi: 10.2196/mhealth.9033] [Medline: 30170994]
31. Pham B, McLeod S. Consonants, vowels and tones across Vietnamese dialects. Int J Speech Lang Pathol 2016 Apr;18(2):122-134. [doi: 10.3109/17549507.2015.1101162] [Medline: 27172848]

Abbreviations

AAPI: Asian Americans and Pacific Islanders
ICAN: International Children Assistance Network
KAB: knowledge-attitude-behavior
LEP: limited English proficiency
NIH: National Institutes of Health
PMI: Precision Medicine Initiative

Edited by C Dias; submitted 30.07.20; peer-reviewed by A Mahnke, B Smith, IV George, K Rosen, F Prazeres, S Prinjha; comments to author 25.09.20; revised version received 19.11.20; accepted 11.12.20; published 11.03.21.

Please cite as:
Ta Park V, Kim A, Cho IH, Nam B, Nguyen K, Vuong Q, Periyakoil VS, Hong YA
Motivation to Participate in Precision Health Research and Acceptability of Texting as a Recruitment and Intervention Strategy Among Vietnamese Americans: Qualitative Study
JMIR Mhealth Uhealth 2021;9(3):e23058
URL: https://mhealth.jmir.org/2021/3/e23058
doi:10.2196/23058
PMID:33704080

Original Paper

A Novel Remote Follow-Up Tool Based on an Instant Messaging/Social Media App for the Management of Patients With Low Anterior Resection Syndrome: Pilot Prospective Self-Control Study

Fan Liu[1,2*], PhD, MD; Peng Guo[1,2*], PhD, MD; Xiangqian Su[3], PhD, MD; Ming Cui[3], PhD, MD; Jianlong Jiang[4], MD; Suo Wang[4], MSc, MD; Zhouman Yu[5], MD; Runhe Zhou[5], MD; Yingjiang Ye[1,2], PhD, MD

[1]Department of Gastroenterological Surgery, Peking University People's Hospital, Beijing, China

[2]Beijing Key Laboratory of Colorectal Cancer Diagnosis and Treatment Research, Beijing, China

[3]Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Department of Gastrointestinal Surgery IV, Peking University Cancer Hospital and Institute, Beijing, China

[4]Department of General Surgery, Changshu Hospital Affiliated to Soochow University, First People's Hospital of Changshu City, Changshu, China

[5]Department of General Surgery, QiLu Hospital (Qingdao), Cheeloo College of Medicine, Shandong University, Qingdao, China

[*]these authors contributed equally

Corresponding Author:
Yingjiang Ye, PhD, MD
Beijing Key Laboratory of Colorectal Cancer Diagnosis and Treatment Research
6A ward, Department of Gastroenterological Surgery, Peking University People's Hospital
No 11 Xizhimen South Street, Xicheng District
Beijing, 100044
China
Phone: 86 10 88326600
Email: yingjiangye@pkuph.edu.cn

Related Article:

This is a corrected version. See correction statement: http://mhealth.jmir.org/2021/4/e29325/

Abstract

Background: Low anterior resection syndrome (LARS) is a common functional disorder that develops after patients with rectal cancer undergo anal preservation surgery. Common approaches to assess the symptoms of patients with LARS are often complex and time-consuming. Instant messaging/social media has great application potential in LARS follow-up, but has been underdeveloped.

Objective: The aim of this study was to compare data between a novel instant messaging/social media follow-up system and a telephone interview in patients with LARS and to analyze the consistency of the instant messaging/social media platform.

Methods: Patients with R0 resectable rectal cancer who accepted several defecation function visits via the instant messaging/social media platform and agreed to a telephone interview after the operation using the same questionnaire including subjective questions and LARS scores were included. Differences between the 2 methods were analyzed in pairs and the diagnostic consistency of instant messaging/social media was calculated based on telephone interview results.

Results: In total, 21 questionnaires from 15 patients were included. The positive rates of defecation dissatisfaction, life restriction, and medication use were 10/21 (48%), 11/21 (52%), and 8/21 (38%) for telephone interview and 10/21 (48%), 13/21 (62%), and 5/21 (24%) for instant messaging/social media, respectively. No statistically significant difference was observed between instant messaging/social media and telephone interview in terms of total LARS score (mean 22.4 [SD 11.9] vs mean 24.7 [SD 10.7], P<.21) and LARS categories (Z=–0.264, P=.79); however, instant messaging/social media showed a more negative tendency. The kappa values of 3 subjective questions were 0.618, 0.430, and 0.674, respectively. The total LARS scores were consistent between both groups (Pearson coefficient 0.760, P<.001; category correlation coefficient 0.570, P=.005). Patients with major

LARS had highly consistent results, with sensitivity, specificity, kappa value, and *P* value of 77.8%, 91.7%, 0.704, and .001, respectively.

Conclusions: Instant messaging/social media can be a major LARS screening method. However, further research on information accuracy and user acceptance is needed before implementing a mature system.

Trial Registration: ClinicalTrials.gov NCT03009747; https://clinicaltrials.gov/ct2/show/NCT03009747

(JMIR Mhealth Uhealth 2021;9(3):e22647) doi:10.2196/22647

KEYWORDS

instant messaging social media; rectal cancer; low anterior resection syndrome; follow-up; telephone interview

Introduction

Low anterior resection syndrome (LARS) is a common functional disorder that develops after patients with rectal cancer undergo anal preservation surgery [1,2]. Its symptoms include changes in defecation frequency, rhythm disorder, incontinence, and constipation, which have been proven to seriously affect the postoperative quality of life [3]. About 30%-55% of patients with rectal cancer have severe LARS symptoms after they complete anal preservation surgery, which can last for several years [4,5].

With improvements in comprehensive treatments for rectal cancer, patients with long-term survival continue to have an increasing demand for LARS treatment, which is a challenge to the current medical system. Its symptoms are variable and persistent, so patients have frequent clinic needs. Common approaches to assess the symptoms of patients with LARS in practice, including face-to-face clinic interviews, post or email questionnaires, and telephone interviews, are often complex and time-consuming, especially when the population of patients with this functional disorder is large [6].

With the popularity of smartphones and mobile internet, remote network technology is changing traditional medical behavior [7,8]. For example, instant messaging/social media represented by WeChat (Tencent Computer System Co., Ltd.) has penetrated Chinese people's daily life. WeChat has evolved into an information exchange platform widely accepted by people because of its high popularity rate and rich expansibility in China. Medical institutions can develop customized small programs to communicate with patients in batches with the help of many third-party software providers. The positive effects of online follow-up developed based on WeChat platforms in chronic disease prevention and multiple kinds of cancers have been confirmed in several studies [9-14].

Technically, using an instant messaging/social media platform to collect follow-up information has the advantages of privacy, speed, user friendliness, economical value, and fragmented time, which are very suitable for the follow-up needs of patients with LARS. However, an advanced follow-up system based on an instant messaging/social media platform has yet to be made available for clinical use among patients with LARS. To fill this gap, the research team developed an instant messaging/social media follow-up system based on WeChat for patients with LARS and designed a prospective self-controlled clinical study. The team also compared the data from instant messaging/social media and telephone interview and analyzed the consistency of the instant messaging/social media platform.

Methods

Study Population

This study was a subproject of the Bas-1611 study. From January 2017 to April 2018, the researchers invited patients who were diagnosed with rectal cancer from 14 medical centers in China and about to receive radical anal preservation surgery. Patients who were histologically proven to have rectal adenocarcinoma 0-12 cm from the anal verge as confirmed by rigid sigmoidoscopy, aged 18 or older, and expected to undergo R0 resection and primary reconstruction were prospectively included. Patients who agreed to participate in the Bas-1611 study and provided informed consent were invited in accordance with the voluntary principle and smartphone usage habits.

Study Registration

As a subproject of Bas-1611, this study was approved by the ethics committee of the competent authority (2017PHB011-01) and registered on the ClinicalTrials.gov website (NCT03009747). All the enrolled patients received a full informed consent document and signed their consent forms.

Study Design

Development of Questionnaires

A set of questionnaires was established in accordance with the standard of patient-reported outcome measures, including patients' survival status, defecation satisfaction survey, and Chinese version of the LARS score scale [15]. The questionnaire was transferred into an online version and integrated on the research management platform developed by Servbus Technology Co., Ltd. After signing the informed consent form, the participants scanned the QR code provided by the researcher with their WeChat app and then completed the registration on the platform. The platform automatically sent the questionnaire to the registered WeChat account 3 (90 days), 6 (180 days), and 12 months (365 days) after the operation, and the feedback was saved on the platform database after the participants completed the questionnaire.

Telephone Interview

Following the Bas-1611 plan, the participants also received telephone interview at 3 time points. Telephone interview was conducted 1-2 weeks after the WeChat client push. The telephone interview questionnaire was identical to the instant

messaging/social media version, but the follow-up personnel would not see the result of the instant messaging/social media follow-up during the interview. All patients enrolled in the study were interviewed by a third-party follow-up team that was employed and trained in a functional follow-up. The interview was properly audio recorded and stored offline to ensure the traceability and quality of research data. The recorded data were rechecked by the experts of the research group against the interview results before data were statistically analyzed to guarantee the accuracy of the telephone interview.

Research Questionnaire

Both instant messaging/social media and telephone interview questionnaires contained exactly the same questions and sequences. The following 3 questionnaire forms were used: (1) patient-reported outcome measures (including patients' survival status and demographic and clinical characteristics); (2) defecation satisfaction survey (including 3 subjective questions Q1, Q2, and Q3); and (3) the Chinese version of the LARS score. The 3 subjective questions included defecation satisfaction, life restriction, and medication use, which required a "yes" or "no" answer. The LARS score was defined as the total score of the items of the questionnaire containing 5 single-choice questions with a corresponding score for each option. Each of the 5 questions tested a single symptom of bowel function, including flatus incontinence, loose stool incontinence, frequency change, clustering, and defecation urgency. The questionnaire could be used to evaluate each patient's defecation function based on the total score (range 0-42 points), which was divided into 3 categories from best to worst: no LARS (0-20), minor LARS (21-29), and major LARS (30-42) [16].

Statistical Analysis

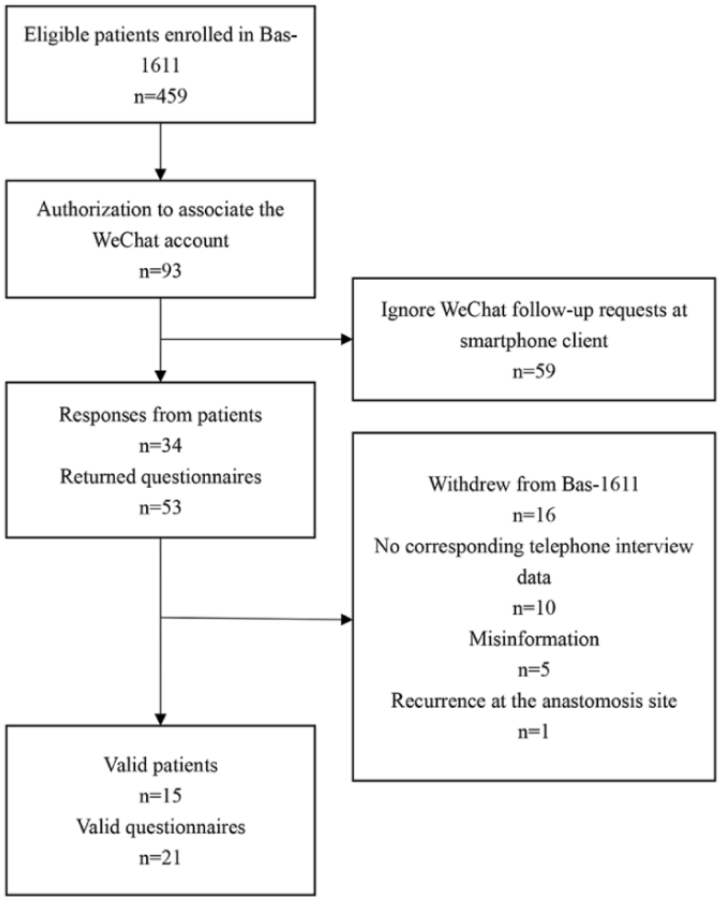
Data were collected in pairs for both types at 1 follow-up node and included in statistical analysis. Pearson correlation analysis and paired Student test were conducted for the correlation of continuous variables. Wilcoxon signed rank test was carried out for categorized and ordered variables, and Kendall tau-b correlation analysis was performed for significant correlation. The diagnostic consistency of the kappa value of the instant messaging/social media follow-up method was calculated on the basis of the telephone interview results. $P<.05$ was considered statistically significant. Analyses were performed using SPSS Statistics version 24.0 (IBM Corp).

Results

Enrolled Patients, Follow-Up, and Drop-Outs

In total, 459 patients from 14 centers were enrolled in Bas-1611 from January 2017 to April 2019. Among them, 93 patients voluntarily received an additional follow-up via the instant messaging/social media platform and successfully registered a WeChat account. During the study, 212 questionnaires were automatically pushed to these patients via the research management platform. A total of 53 instant messaging/social media questionnaires from 34 patients were collected. Among them, 16 patients dropped out of the Bas-1611 study because of protocol deviation, 10 instant messaging/social media questionnaires had no corresponding telephone interview record, and 5 questionnaires were mistakenly filled before stoma reversal. One patient dropped out because of tumor recurrence. A total of 21 paired instant messaging/social media questionnaires from 15 patients were eventually included in the final analysis (Figure 1).

Figure 1. Flowchart of 15 patients.



The 15 patients comprising 10 males and 5 females with an average age of 60.6 (SD 8.3) years were from 4 research centers participating in Bas-1611. A total of 3 patients received preoperative neoadjuvant radiotherapy (cases 3, 5, and 12), while 6 patients received temporary stoma (cases 2, 3, 5, 7, 12, 14) and completed reversion surgery before the expiration date (Table 1).

Table 1. Enrolled patient characteristics (N=15).

Variable type and category	Values
Demographic characteristic	
Male, n (%)	10 (67)
Female, n (%)	5 (33)
Age at time of surgery (years), mean (SD), range	60.6 (8.3), 49-74
BMI (kg/m^2), mean (SD), range	24.5 (1.9), 22.1-27.8
Distance to anal verge (cm), mean (SD), range	7.3 (2.3), 4-12
Preoperative TNM classification, n (%)	
0	1 (7)
I	4 (27)
II	5 (33)
III	4 (27)
IV	1 (7)
Neoadjuvant radiotherapy, n (%)	
Yes	3 (20.0)
No	12 (80.0)
Diverting stoma, n (%)	
Yes	6 (40.0)
No	9 (60.0)

In Table 2, the answers to the subjective part of 11 questionnaires were completely consistent; 3 questionnaire responses (from case 7, case 13, and case 15) had 2 inconsistent answers, and the remaining 7 questionnaire responses each had 1 inconsistent answer. The results of the telephone interview indicated that 10/21 (48%), 11/21 (52%), and 8/21 (38%) of the participants reported dissatisfaction with their defecation, the effect on the quality of life, and medication use for bowel symptoms, respectively. The corresponding results for the 3 options in the instant messaging/social media follow-up were 10/21 (48%), 13/21 (62%), and 5/21 (24%).

Table 2. Response to the 3 subjective questions.

Case number	Time node (months)	Q1[a]		Q2[b]		Q3[c]	
		IMSM[d]	TI[e]	IMSM	TI	IMSM	TI
1	3	Yes	Yes	No	No	No	No
1	6	Yes	Yes	No	No	No	No
2	12	Yes	Yes	No	Yes	No	No
3	12	No	No	Yes	Yes	No	No
4	3	No	No	Yes	Yes	Yes	Yes
5	6	No	No	No	No	No	No
6	6	No	No	Yes	Yes	Yes	Yes
7	6	Yes	Yes	Yes	Yes	No	No
7	12	Yes	No	No	Yes	No	No
8	12	Yes	No	No	No	No	No
9	6	No	No	Yes	Yes	No	No
10	6	Yes	Yes	Yes	No	No	No
11	3	No	No	Yes	Yes	No	Yes
12	6	No	Yes	Yes	Yes	No	No
13	3	No	No	Yes	Yes	Yes	Yes
13	6	No	No	Yes	Yes	No	Yes
13	12	Yes	Yes	Yes	No	No	Yes
14	12	Yes	Yes	No	No	No	No
15	3	Yes	Yes	Yes	No	Yes	Yes
15	6	No	Yes	Yes	No	Yes	Yes
15	12	Yes	Yes	No	No	No	No

[a]Q1: Are you satisfied with the current bowel function?

[b]Q2: Does current defecation affect your daily life?

[c]Q3: Do you use medication to improve your defecation?

[d]IMSM: instant messaging/social media.

[e]TI: telephone interview.

Multimedia Appendix 1 provides the LARS score results of 21 paired questionnaires. The total score of the LARS questionnaire obtained using the 2 follow-up methods was the same, and ranged from 0 to 39. Although the average score of the instant messaging/social media group was relatively high, the paired *t* test revealed that the 2 methods had no statistically significant difference (22.4 [SD 11.9] versus 24.7 [SD 10.7], t_{20}=1.285, *P*=.21). The total score of 10/21 questionnaires (48%) was higher than that of the telephone follow-up, the scores of 7/21 questionnaires (33%) were consistent, and the scores of 4/21 questionnaires (19%) were lower than those of the telephone follow-up (Table 3 and Figure 2).

Table 3. Statistical analysis of the LARS score response.[a]

Variables/responses	Score range	Consistency, n	IMSM[b] score high, n	TI[c] score high, n	Nonparametric test		Consistency check	
					Wilcoxon *Z* value	*P* value	Kappa value	*P* value
LARS[d] score	0-42	7	10	4	−1.509	.13	0.760[a]	<.001[a]
LARS category	—	14	4	3	−0.264	.79	0.490	.001
Q1	0-7	10	9	2	−2.032	.04	0.206	.12
Q2	0-3	15	6	0	2.449	.01	0.523	<.001
Q3	0-5	12	0	8	−2.555	.01	0.472	<.001
Q4	0-11	15	4	2	−0.638	.52	0.543	<.001
Q5	0-16	18	1	2	0.001	>.99	0.786	<.001

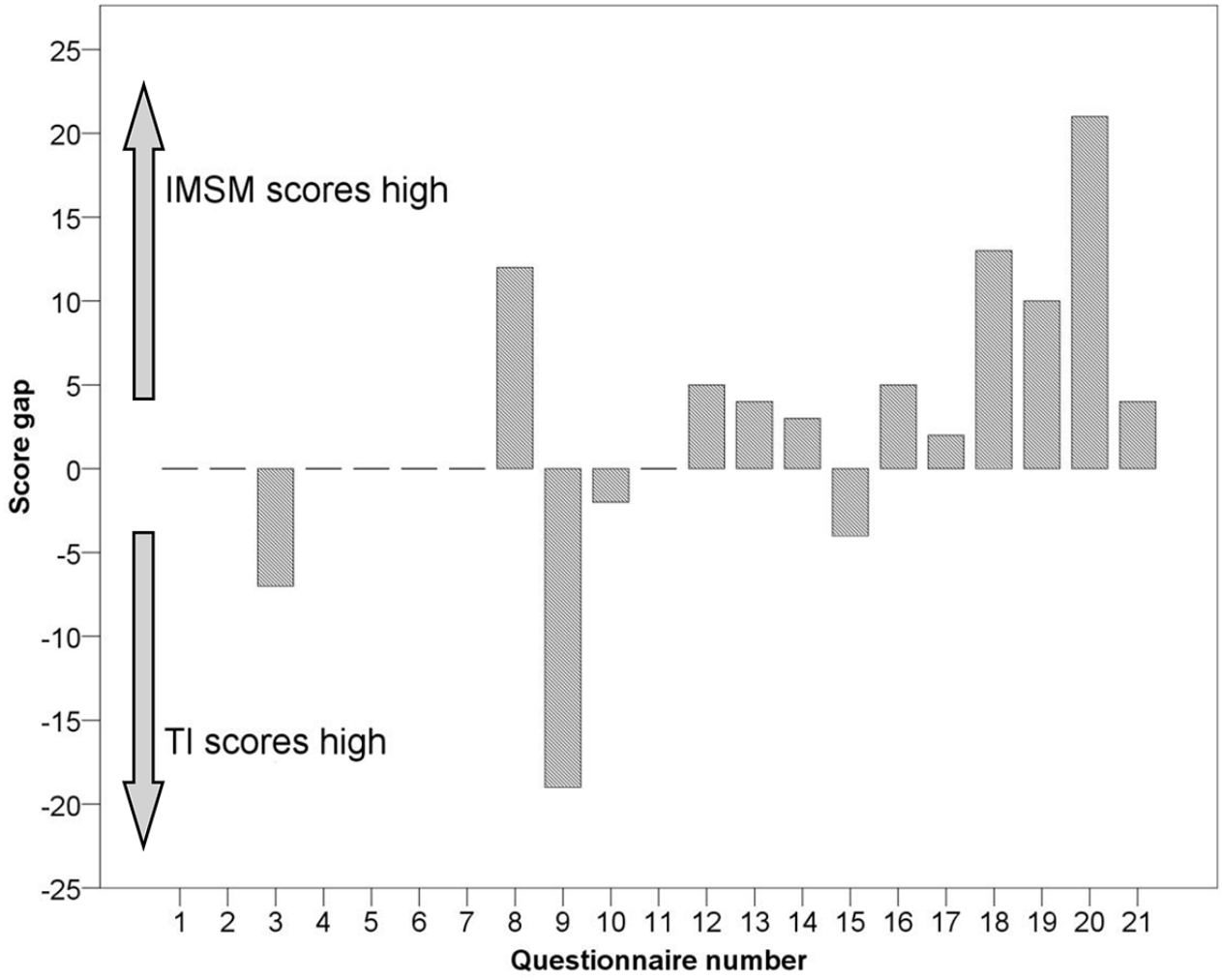
[a]Because LARS score is a continuous variable, the Pearson coefficient and corresponding *P* value are calculated here.

[b]IMSM: instant messaging/social media.

[c]TI: telephone interview.

[d]LARS: low anterior resection syndrome.

Figure 2. Bar chart of the LARS score difference of 21 paired questionnaires. IMSM: instant messaging/social media; LARS: low anterior resection syndrome; TI: telephone interview.



The proportions of no LARS, minor LARS, and major LARS in the telephone interview group were 9, 3, and 9, respectively, whereas their corresponding proportions in the instant messaging/social media interview were 7, 6, and 8. Inconsistencies were observed in the LARS categories of 7 questionnaires, and 2 cases (cases 7 and 15) had extreme inconsistencies. Although the proportion of minor LARS was higher in the instant messaging/social media outcome than in

the telephone interview group, the difference was not significant (Z=–0.264, P=.79).

Each symptom was further analyzed. The instant messaging/social media groups performed worse in 4 symptoms but not in the frequency of bowel movement (Q3). Among them, flatus incontinence (Q1; P=.04) and loose stool incontinence (Q2; P=.01) showed a statistically significant difference in the Wilcoxon signed rank test. The frequency of bowel movement in the instant messaging/social media questionnaire was significantly better (Q3, P=.01). No significant difference was observed in the results of "Q4: clustering"(P=.52) and "Q5: defecation urgency"(P>.99) Although Q1, Q2, and Q3 were significantly different, the influence on the total score was masked by the high score of Q4 and Q5.

Severe deviations were observed in cases 7 and 15 in the subjective questionnaire and in the LARS score. Case 7, a 56-year-old man with an eye condition, asked his wife to make the choice on the WeChat client for him. Her judgment differed greatly from the patient's actual feelings, even though they live together. Case 15 was a 50-year-old woman who had repeated episodes of intestinal infection after surgery, which resulted in recurrent diarrhea and incontinence. During the instant messaging/social media follow-up, she described the most recent and severe defecation symptoms, but the symptoms caused by the infection were difficult to be distinguished. Conversely, our telephone interview follow-up staff accurately determined her true defecation condition in the noninfected state.

The patients reported more negative functional results on the instant messaging/social media platform, although no statistically significant difference was obtained due to sample size limitations (Z=–0.264, P=.79). For example, 10 (71%) of the 14 questionnaires with inconsistent LARS scores were high, and 4 (67%) of the 6 questionnaires with inconsistent quality of life evaluations chose "yes." This trend was still evident after cases 7 and 15 were removed.

Diagnostic Consistency Analysis

The kappa values of the 3 subjective questions were statistically acceptable (Table 4). The sensitivity and specificity of the questionnaire satisfaction survey were 80.0% and 81.8%, respectively, and the diagnostic consistency coefficient was 0.618 (P=.006). The results of the quality of life survey were relatively poor; that is, the sensitivity and specificity were 81.8% and 60.0%, respectively. The diagnostic consistency coefficient was 0.430 (P=.04). Three questionnaire responses from cases 11 and 13 with differences on medication use were false negative, with a sensitivity of 62.5%, a specificity of 100%, and a diagnostic consistency coefficient of 0.674 (P=.001). They were found to have taken an over-the-counter Chinese herbal medicine, which can improve defecation, during the telephone follow-up, and their decision to use medicines was confirmed as a "yes" after the panel discussion.

Table 4. Consistency test of the 3 subjective survey results in the instant messaging/social media and telephone interview questionnaires.

Questionnaire	Consistent, n	Inconsistent, n	Sensitivity, %	Specificity, %	Kappa value	*P* value
Q1	17	4	80.0	81.8	0.618	.006
Q2	15	6	81.8	60.0	0.430	.049
Q3	18	3	62.5	100	0.674	.001

The results of the total LARS score were consistent between the 2 groups. The Pearson coefficient was 0.760 (P<.001), while the category correlation coefficient was 0.570 (P=.005). The kappa value for diagnostic consistency of the LARS category was 0.490 (P=.001). Among the 3 categories, the consistency of major LARS was the best, with a sensitivity of 77.8% and a specificity of 91.7% (κ=0.704, P=.001). Urgency was the most consistent in diagnosis among all the symptoms (κ=0.786), with only 3 questionnaire responses having different results. Conversely, no statistically significant kappa value was obtained for flatus incontinence (κ=0.206, P=.12).

Discussion

Principal Results

In this study, a follow-up system for patients with LARS was established on the basis of the instant messaging/social media platform (WeChat app). The results obtained with this method were paired with those obtained with traditional telephone interview. Our findings indicated that the functional outcome of the instant messaging/social media platform was basically consistent with that of telephone interview. In particular, patients with major LARS had a strong consistency and showed more negative functional evaluation trends in the instant messaging/social media platform.

Application Trend of Instant Messaging/Social Media in Functional Follow-Up

Combined with physical examination and treatment, a face-to-face clinical interview is the most effective way of a functional follow-up [8]. However, it takes time and entails labor costs; furthermore, doctors with LARS management experience are lacking in China. Therefore, LARS management is impossible to be arranged postoperatively for every patient with rectal cancer [17]. In this study, instead of a face-to-face follow-up, telephone interview was used as the control setting, which was also based on the current situation of insufficient outpatient resources for diseases related to defecation function. New methods, such as telephone and remote follow-up, are vital complements to face-to-face follow-up, and their better patient satisfaction and lower cost have been confirmed in the follow-up of patients with cancer and other functional diseases [18,19].

The trend to turn to social media among doctors and patients with cancer for an interchange of disease information is growing [20,21]. However, Pellino et al [20] found that the knowledge acquired by patients with colorectal cancer from open social

media is mixed and varied; authoritative arguments are also lacking, while useful knowledge is often overwhelmed by the mass of information. Given the widespread influence of social media, researchers should use it to issue recruitment notices [3,12], but discussing specific symptoms in open social media is difficult, especially for a very private functional disorder such as LARS.

Consistency of the Instant Messaging/Social Media Platform

Instant messaging/social media has potential as a follow-up platform of patients with LARS because of its privacy, security, convenience, and wide coverage nature, and its accuracy is supported by our research evidence. In this study, responses of patients with major LARS had high consistency, possibly because the symptoms of major LARS are hard to allay. By contrast, patients with mild LARS sometimes had no symptoms; therefore, their feedback fluctuated. This may be because telephone interview was conducted 1-2 weeks after the WeChat client push. The screening and treatment of patients with major LARS are a key part of LARS management, and such patients have difficulty obtaining adequate help from a general oncology client. The stable performance of our follow-up system in major LARS makes it suitable for the follow-up and evaluation of patients with major LARS.

Patients who have cancer and independently complete the questionnaire tend to overstate the extremes on the quality of life [22], especially those whose long-term survival is no longer threatened by cancer. This study explained the higher scores on the instant messaging/social media platform, and they are common in similar studies [23]. The moderate exaggeration of negative feelings in patients is emotionally understandable and may even be common, but whether such exaggeration affects the accuracy of follow-up remains to be further studied.

One of the advantages of LARS scoring is that its logic of question is simple and clear, so patients can easily choose a response [16]. However, such a concise description can likely lead to misunderstanding in the instant messaging/social media platform. In this study, some patients misinterpreted incontinence in flatus and the average number of bowel movement per day. This misunderstanding led to a large discrepancy in Q1 and Q3. A previous study [24] indicated that the LARS questionnaire has defects in evaluating symptoms such as emptying disorder. A functional follow-up on smartphones is a special application scenario in which no professional guidance is available, and appropriate adjustments should be made based on the patient population studied; for example, explanatory words for easily confused parts should be added.

Patients' Willingness and Satisfaction

Patients' willingness to use the instant messaging/social media platform also depends on whether this new method is more convenient and economical than traditional ones. Against the background of generally improved prognosis of colorectal cancer, the medical system is barely being maintained, and providing satisfactory follow-up for outpatient services is difficult. Dai et al [11] found that difficulties in visiting a central medical institution prompt 66.1% of patients to use social media for tracking and feedback. Teagle et al [8] believed that using a remote follow-up technology is an economically feasible solution, which can effectively reduce the burden of the follow-up personnel and reduce the travel cost and missed days of work. Smartphones and instant messaging/social media may be a barrier for some elderly patients, but given that the rectal cancer morbidity has a youth-oriented tendency, this technology may be accepted by more patients in the future. In this study, 10 of the 53 returned questionnaires (19%) were followed up without the corresponding telephone interview, indicating that the convenience of instant messaging/social media might further improve patients' follow-up intention and response rate in the future.

Limitations

The main limitation of this study was the low response rate in the instant messaging/social media group, which lead to small sample size in the final analysis. The reasons for low response may be as follows: (1) Elderly patients with rectal cancer were not active users of smartphones; (2) We underestimated the huge information of WeChat app. According to the protocol, we sent only 1 follow-up message at each follow-up node, while the WeChat app may receive dozens or even hundreds of messages every day, due to which some patients failed to notice the follow-up reminder; and (3) The design of the Bas-1611 study made telephone interview available to all patients regardless of their response to the instant messaging/social media follow-up request, possibly resulting in an excessively low WeChat response rate.

Selective bias and outlier results are inevitable because of the small sample size of this study. For the patients to be proficient in using WeChat, younger or better educated patients are needed to be enrolled in studies of this kind. The complexity of instant messaging/social media user behavior leads to the inaccuracy of follow-up information (such as cases 7 and 15), which still needs to be solved in a follow-up study. Inspired by our work, a new randomized controlled trial (BaS-1904, NCT03669237) is committed to further explore the issues of LARS patient management, and it is expected that the aforesaid limitations will also be improved in the new study.

Conclusions

The instant messaging/social media system provides a promising solution to accommodate the primary follow-up needs of patients with LARS by integrating complex functional follow-up tools into smartphone apps. Although it is currently not a substitute for manual follow-up, it has the potential of becoming a major LARS screening method. However, further research on response rate, information accuracy, and user acceptance is needed before an advanced system can be implemented.

Acknowledgments

This project is supported by the National Key R&D Program of China (Grant No. 2017YF0908203). The authors thank Mr Tao Shao from the Servbus Technology Co., Ltd. (Beijing, China) for his support in this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Response of LARS Score.
[DOCX File , 24 KB - mhealth_v9i3e22647_app1.docx]

References

1. van Heinsbergen M, den Haan N, Maaskant-Braat AJ, Melenhorst J, Belgers EH, Leijtens JW, et al. Functional bowel complaints and quality of life after surgery for colon cancer: prevalence and predictive factors. Colorectal Dis 2020 Feb;22(2):136-145. [doi: 10.1111/codi.14818] [Medline: 31397962]
2. Pieniowski EH, Palmer GJ, Juul T, Lagergren P, Johar A, Emmertsen KJ, et al. Low Anterior Resection Syndrome and Quality of Life After Sphincter-Sparing Rectal Cancer Surgery: A Long-term Longitudinal Follow-up. Dis Colon Rectum 2019 Jan;62(1):14-20. [doi: 10.1097/DCR.0000000000001228] [Medline: 30394987]
3. Keane C, Fearnhead NS, Bordeianou LG, Christensen P, Basany EE, Laurberg S, LARS International Collaborative Group. International Consensus Definition of Low Anterior Resection Syndrome. Dis Colon Rectum 2020 Mar;63(3):274-284 [FREE Full text] [doi: 10.1097/DCR.0000000000001583] [Medline: 32032141]
4. Sterner A, Derwinger K, Staff C, Nilsson H, Angenete E. Quality of life in patients treated for anal carcinoma-a systematic literature review. Int J Colorectal Dis 2019 Sep;34(9):1517-1528. [doi: 10.1007/s00384-019-03342-x] [Medline: 31324957]
5. Kverneng Hultberg D, Svensson J, Jutesten H, Rutegård J, Matthiessen P, Lydrup M, et al. The Impact of Anastomotic Leakage on Long-term Function After Anterior Resection for Rectal Cancer. Dis Colon Rectum 2020 May;63(5):619-628. [doi: 10.1097/DCR.0000000000001613] [Medline: 32032197]
6. Beaver K, Wilson C, Procter D, Sheridan J, Towers G, Heath J, et al. Colorectal cancer follow-up: patient satisfaction and amenability to telephone after care. Eur J Oncol Nurs 2011 Feb;15(1):23-30. [doi: 10.1016/j.ejon.2010.05.006] [Medline: 20584628]
7. Garfinkle R, Wong-Chong N, Petrucci A, Sylla P, Wexner SD, Bhatnagar S, et al. Assessing the readability, quality and accuracy of online health information for patients with low anterior resection syndrome following surgery for rectal cancer. Colorectal Dis 2019 May;21(5):523-531. [doi: 10.1111/codi.14548] [Medline: 30609222]
8. Teagle A, Gilbert D. Remote follow-up strategies after cancer treatment: a lot of opportunities. Clin Oncol (R Coll Radiol) 2014 Oct;26(10):622-624. [doi: 10.1016/j.clon.2014.05.009] [Medline: 24969682]
9. Dorje T, Zhao G, Scheer A, Tsokey L, Wang J, Chen Y, et al. SMARTphone and social media-based Cardiac Rehabilitation and Secondary Prevention (SMART-CR/SP) for patients with coronary heart disease in China: a randomised controlled trial protocol. BMJ Open 2018 Jun 30;8(6):e021908 [FREE Full text] [doi: 10.1136/bmjopen-2018-021908] [Medline: 29961032]
10. Trigo J, Eguzkiza A, Martinez-Espronceda M, Serrano L. A cardiovascular patient follow-up system using Twitter and HL7. 2013 Sep Presented at: Computing in Cardiology; September 22nd – 25th, 2013; Zaragoza, Spain p. 33-36 URL: https://www.cinc.org/archives/2013/pdf/0033.pdf
11. Dai M, Xu J, Lin J, Wang Z, Huang W, Huang J. Willingness to Use Mobile Health in Glaucoma Patients. Telemed J E Health 2017 Dec;23(10):822-827. [doi: 10.1089/tmj.2016.0254] [Medline: 28418773]
12. Brittain K, Pennings Kamp KJ, Salaysay Z. Colorectal Cancer Awareness for Women via Facebook: A Pilot Study. Gastroenterol Nurs 2018;41(1):14-18 [FREE Full text] [doi: 10.1097/SGA.0000000000000294] [Medline: 29373351]
13. Sheng J, Zhao Z, Zhang H, Huang R. Effect of Internet-based follow-up on antiviral treatment compliance and quality of life in discharged patients with chronic hepatitis B. WCJD 2017 Nov 8;25(31):2813-2818. [doi: 10.11569/wcjd.v25.i31.2813]
14. Ye X, Wei J, Li Z, Niu X, Wang J, Chen Y, et al. Design and implementation of a mobile system for lung cancer patient follow-up in China and initial report of the ongoing patient registry. Oncotarget 2017 Jan 17;8(3):5487-5497 [FREE Full text] [doi: 10.18632/oncotarget.13720] [Medline: 27911868]
15. Hou X, Pang D, Lu Q, Yang P, Jin S, Zhou Y, et al. Validation of the Chinese version of the low anterior resection syndrome score for measuring bowel dysfunction after sphincter-preserving surgery among rectal cancer patients. Eur J Oncol Nurs 2015 Oct;19(5):495-501. [doi: 10.1016/j.ejon.2015.02.009] [Medline: 25813530]
16. Emmertsen KJ, Laurberg S. Low anterior resection syndrome score: development and validation of a symptom-based scoring system for bowel dysfunction after low anterior resection for rectal cancer. Ann Surg 2012 May;255(5):922-928. [doi: 10.1097/SLA.0b013e31824f1c21] [Medline: 22504191]

17. Strand E, Nygren I, Bergkvist L, Smedh K. Nurse or surgeon follow-up after rectal cancer: a randomized trial. Colorectal Dis 2011 Sep;13(9):999-1003. [doi: 10.1111/j.1463-1318.2010.02317.x] [Medline: 20478003]
18. Dixon P, Beaver K, Williamson S, Sutton C, Martin-Hirsch P, Hollingworth W. Cost-Consequence Analysis Alongside a Randomised Controlled Trial of Hospital Versus Telephone Follow-Up after Treatment for Endometrial Cancer. Appl Health Econ Health Policy 2018 Jun;16(3):415-427 [FREE Full text] [doi: 10.1007/s40258-018-0378-6] [Medline: 29651778]
19. Everitt HA, Landau S, O'Reilly G, Sibelli A, Hughes S, Windgassen S, ACTIB trial group. Assessing telephone-delivered cognitive-behavioural therapy (CBT) and web-delivered CBT versus treatment as usual in irritable bowel syndrome (ACTIB): a multicentre randomised trial. Gut 2019 Sep;68(9):1613-1623 [FREE Full text] [doi: 10.1136/gutjnl-2018-317805] [Medline: 30971419]
20. Pellino G, Simillis C, Qiu S, Rasheed S, Mills S, Warren O, et al. Social media and colorectal cancer: A systematic review of available resources. PLoS One 2017;12(8):e0183031 [FREE Full text] [doi: 10.1371/journal.pone.0183031] [Medline: 28832603]
21. Tsuya A, Sugawara Y, Tanaka A, Narimatsu H. Do cancer patients tweet? Examining the twitter use of cancer patients in Japan. J Med Internet Res 2014 May 27;16(5):e137 [FREE Full text] [doi: 10.2196/jmir.3298] [Medline: 24867458]
22. Mercier M, Schraub S, Bransfield D, Barthod L. [Measurement of quality of life. Application to the screening of psychological distress in cancer patients]. Bull Cancer 1992;79(2):193-204. [Medline: 1392158]
23. Corner J, Wagland R, Glaser A, Richards SM. Qualitative analysis of patients' feedback from a PROMs survey of cancer patients in England. BMJ Open 2013;3(4):e002316 [FREE Full text] [doi: 10.1136/bmjopen-2012-002316] [Medline: 23578681]
24. Ribas Y, Aguilar F, Jovell-Fernández E, Cayetano L, Navarro-Luna A, Muñoz-Duyos A. Clinical application of the LARS score: results from a pilot study. Int J Colorectal Dis 2017 Mar;32(3):409-418. [doi: 10.1007/s00384-016-2690-7] [Medline: 27796496]

Abbreviations

LARS: low anterior resection syndrome

Edited by G Eysenbach; submitted 19.07.20; peer-reviewed by E Sutsunbuloglu, K Aguirre, R Vilela; comments to author 26.11.20; revised version received 13.01.21; accepted 30.01.21; published 19.03.21.

Please cite as:
Liu F, Guo P, Su X, Cui M, Jiang J, Wang S, Yu Z, Zhou R, Ye Y
A Novel Remote Follow-Up Tool Based on an Instant Messaging/Social Media App for the Management of Patients With Low Anterior Resection Syndrome: Pilot Prospective Self-Control Study
JMIR Mhealth Uhealth 2021;9(3):e22647
URL: https://mhealth.jmir.org/2021/3/e22647
doi:10.2196/22647
PMID:33739295

Original Paper

Going Remote—Demonstration and Evaluation of Remote Technology Delivery and Usability Assessment With Older Adults: Survey Study

Jordan R Hill[1], MSIE, PhD; Addison B Harrington[2], BS; Philip Adeoye[2], MS; Noll L Campbell[2,3], MS, PharmD; Richard J Holden[1,2], PhD

[1]Department of Medicine, Indiana University School of Medicine, Indianapolis, IN, United States

[2]Center for Aging Research, Regenstrief Institute, Indianapolis, IN, United States

[3]Department of Pharmacy Practice, College of Pharmacy, Purdue University, West Lafayette, IN, United States

Corresponding Author:
Jordan R Hill, MSIE, PhD
Department of Medicine
Indiana University School of Medicine
1101 W 10th St
Indianapolis, IN, 46202
United States
Phone: 1 7655438559
Email: jrh6@iu.edu

Abstract

Background: The COVID-19 pandemic necessitated "going remote" with the delivery, support, and assessment of a study intervention targeting older adults enrolled in a clinical trial. While remotely delivering and assessing technology is not new, there are few methods available in the literature that are proven to be effective with diverse populations, and none for older adults specifically. Older adults comprise a diverse population, including in terms of their experience with and access to technology, making this a challenging endeavor.

Objective: Our objective was to remotely deliver and conduct usability testing for a mobile health (mHealth) technology intervention for older adult participants enrolled in a clinical trial of the technology. This paper describes the methodology used, its successes, and its limitations.

Methods: We developed a conceptual model for remote operations, called the Framework for Agile and Remote Operations (FAR Ops), that combined the general requirements for spaceflight operations with Agile project management processes to quickly respond to this challenge. Using this framework, we iteratively created *care packages* that differed in their contents based on participant needs and were sent to study participants to deliver the study intervention—a medication management app—and assess its usability. Usability data were collected using the System Usability Scale (SUS) and a novel usability questionnaire developed to collect more in-depth data.

Results: In the first 6 months of the project, we successfully delivered 21 *care packages*. We successfully designed and deployed a minimum viable product in less than 6 weeks, generally maintained a 2-week sprint cycle, and achieved a 40% to 50% return rate for both usability assessment instruments. We hypothesize that lack of engagement due to the pandemic and our use of asynchronous communication channels contributed to the return rate of usability assessments being lower than desired. We also provide general recommendations for performing remote usability testing with diverse populations based on the results of our work, including implementing screen sharing capabilities when possible, and determining participant preference for phone or email communications.

Conclusions: The FAR Ops model allowed our team to adopt remote operations for our mHealth trial in response to interruptions from the COVID-19 pandemic. This approach can be useful for other research or practice-based projects under similar circumstances or to improve efficiency, cost, effectiveness, and participant diversity in general. In addition to offering a replicable approach, this paper tells the often-untold story of practical challenges faced by mHealth projects and practical strategies used to address them.

Trial Registration: ClinicalTrials.gov NCT04121858; https://clinicaltrials.gov/ct2/show/NCT04121858

(JMIR Mhealth Uhealth 2021;9(3):e26702) doi:10.2196/26702

KEYWORDS

COVID-19; mobile usability testing; usability inspection; methods; aging; agile; mobile phone

Introduction

Overview

Technological advances have made it more commonplace and easier for individuals to complete tasks remotely, from online banking to visiting a doctor to paid work from home [1]. The COVID-19 pandemic revealed that remote interaction is also sometimes necessary. The need to "go remote" during the pandemic likely both accelerated innovation in telecommunication and normalized available but underused internet-based services and products, such as telehealth [2], mental health apps [3], and e-commerce [4].

The pandemic also disrupted research, forcing projects to suspend operations or implement remote noncontact methods [5-7]. Paradoxically, some studies of mobile technology faced difficulties converting to entirely remote technology delivery and usability assessment, revealing opportunities for technology projects to experiment with remote operations [8].

This paper reports our experiences in "going remote" to rapidly restart a suspended clinical trial of a mobile health (mHealth) app with older adults. We present our protocol changes, Agile project management approach, and findings from this restart project. Our experiences elucidate the opportunities and challenges of conducting technology research remotely and may inform other attempts at rapid operational change.

Background

Remote delivery and usability assessment of technology is not new [9,10] but have heretofore been largely used with younger, highly educated, technologically experienced, and more homogeneous individuals, for example, college students [9,11] or those using the technology as a part of their employment [12-14]. Only one study was found to employ remote methods with older adults and these were supplemented with in-person sessions for initial training and instruction [15].

Nevertheless, existing remote methods provide a foundation for technology development projects with older, disabled, or other diverse populations in which experience is limited. Potentially useful existing methods to access and evaluate technology include online questionnaires, online forums, self-kept diaries, email, telephone calls, and postal delivery [11,15-17]. More recently, videoconferencing tools are used to replicate in-person, moderated usability testing [8,11,17].

These methods have, to our knowledge, not been adapted for fully remote projects with older adult users, although some work has been done with younger and older adults with various disabilities [18]. This gap is relevant because older adults are the fastest growing population of technology users [19] but also one of the most diverse populations in terms of their motivations to use technology and their physical, cognitive, and sensory capabilities [20]. Additionally, among older adults, factors such as race and ethnicity, income, literacy, education level, age group (eg, <75 vs ≥75 years of age), and community characteristics (eg, rural vs urban) significantly impact access and use of the internet and other technologies [21-26]. Thus, one challenge of "going remote" with older adults is to accommodate between-age and within-age diversity.

Accordingly, prior research has suggested best practices and adjustments for in-person technology assessment with older adults. For example, because think-aloud procedures and psychometric scales (eg, Likert) may not work well for this population, it may be better to engage older adults one-on-one in a guided interview to elicit information [27,28]. Others recommend personalizing usability testing protocols with attention to fidelity in those with cognitive decline or disabilities [29].

Objective

The objective of the reported project, Project C (COVID), was *to remotely deliver and conduct usability testing for an mHealth technology intervention for older adult participants enrolled in a clinical trial of the technology*. The need for remote delivery arose from restrictions on in-person research activities due to a global pandemic and an ethical obligation to reduce risk for older adult participants and study personnel. Meeting the objective required special effort due to participants' diverse technology ownership, experience, motivations, and skill levels. For example, a subset of participants did not have a smartphone or home internet and required the study to provide these. Thus, the team could not simply provide an online link to the app for all participants to download. Below, we describe how we developed these remote operations to deliver and test mHealth technology to a heterogenous older adult population participating in human subjects research during a pandemic.

Methods

Overview

Project C occurred in the context of the Brain Safe randomized clinical trial, described in depth elsewhere [30]. The study enrolled adults aged 60 years or older in Indiana, USA, who were using medications with anticholinergic effects that may increase risk to brain health for older adults. In the trial's parallel treatment arms, participants received a mobile app to use for 12 months, either the Brain Safe intervention app or the Med Safe attention control app. Both were native Android or iOS apps, Health Insurance Portability and Accountability Act (HIPAA) compliant, and optimized for smartphones, with the following features:

1. Self-registration.
2. Passwordless log-in, via verification token for each log-in.
3. Four-tab navigation.
4. A medication list created by the user through searching and browsing, selecting medication choices, and numeric data entry.

5. Monthly reminders to complete medication review.
6. Text and video educational materials.
7. PDF report output and sharing.
8. A score calculator.

Whereas the score calculator in the Med Safe app merely calculated the total number of medications, the Brain Safe app calculated a brain harm risk score, let users simulate the effect on risk of adding or removing anticholinergic medications, showed alternative treatments, and helped users start a conversation with their health care professionals about medication safety. Additional detail on features, design, development, and prior testing are available elsewhere [30,31].

Both apps were designed to be loaded on a smartphone meeting minimum hardware, software, and connectivity requirements. Originally, standard operating procedures in the trial were to conduct installation, account registration, training, first-time use, initial technical support, troubleshooting, and usability observations in person at the participant's home or in a research office, with continuing opportunities for remote or in-person support. A smartphone, typically a 5.8-inch Samsung Galaxy S9 or S10, with voice calling and unlimited text and data plans was provided to any participant needing it for the study's duration.

Project C was launched on March 13, 2020, when in-person trial operations were suspended due to COVID-19-related institutional policies, city and state governmental mandates, and an ethical imperative to reduce risk to participants, study personnel, and the community. At the time, the study had consented and enrolled 56 participants out of a planned 700. Of those enrolled, 7 (13%) had received the Brain Safe or Med Safe app and the remaining 49 (88%) were the target group for which Project C was designed. Over time, Project C provided remote capability for newly enrolled participants and is still ongoing. We present results from the first 6 months of iterative development of Project C, through September 13, 2020.

All procedures in the Brain Safe trial, which was registered at ClinicalTrials.gov (NCT04121858), were approved by the Indiana University Institutional Review Board (IRB). Participants consented to the study, including receiving the app, app-related interactions, and usage data collection, and completed usability and technology acceptance questionnaires. Additional procedures for Project C, for example, providing remote technical support or collecting additional feedback on usability, were conducted for the purpose of improving operations and quality; they were, therefore, not deemed human subjects research. However, changes to allow remote consenting processes with newly enrolled participants were approved as protocol amendments by the IRB.

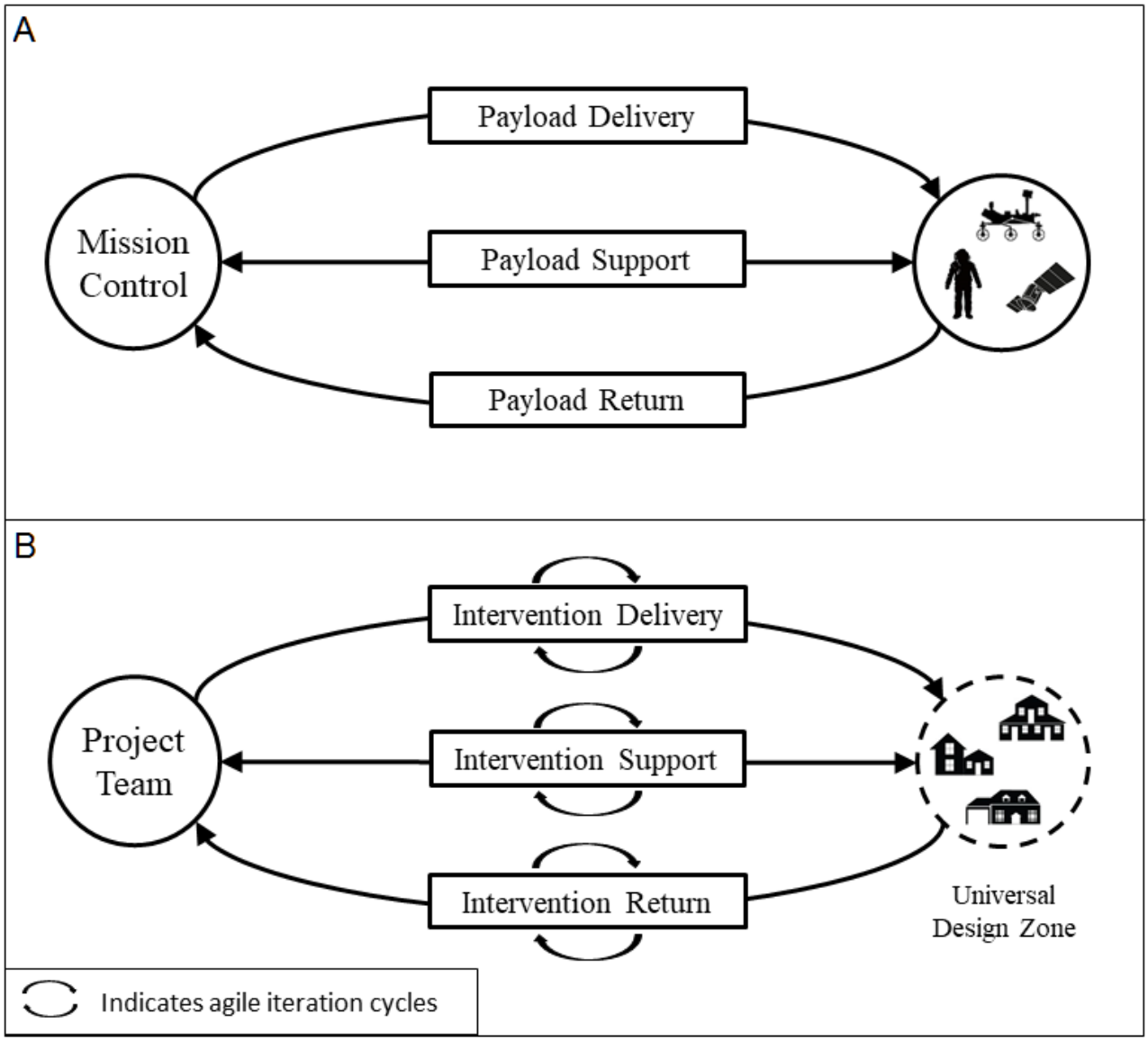
Conceptual Model

We framed the shift to remote operations as a challenge akin to, but less extreme than, spaceflight. Spaceflight missions send people and payloads (eg, satellites, probes, and rovers) to extremely distant locations, including Mars and the International Space Station [32-34]. Mission success requires solving three challenges: (1) delivery (ie, getting payload to its destination), (2) support (ie, remotely identifying and solving problems throughout the mission), and (3) return (ie, retrieving materials or information during or after the mission) (see Figure 1, A) [35-38]. Similarly, the challenge of remote technology intervention research is three-fold: (1) delivery of the technology to a diverse set of distant recipients; (2) support of the technology and its users, including training, technical support, and communication; and (3) return of data (eg, usability assessments), equipment, and feedback regarding status (see Figure 1, B).

To this spaceflight operations model, we added an Agile project management approach [39] to address the need for innovation and timeliness; the Agile approach is represented by feedback loops in Figure 1, B. The Agile approach was introduced by software developers to promote iterative design, delivering products on a short timeline, responding to user feedback, and learning from failures [40]. Agile approaches have since been used to improve health care delivery, for example, to identify, adapt, implement, and evaluate evidence-based clinical services [41]. Holden and Boustani [42] argued that replacing status quo health care operations with an Agile Mindset could help organizations more nimbly and effectively cope with crises, including the COVID-19 pandemic. In our work, we adopted the three principles of an Agile Mindset proposed by those authors:

1. Sprints. Accelerate progress toward a minimum viable product, to be progressively tested and improved in sprints.
2. Sensors. Embed sensors to collect timely, nonjudgmental, and actionable feedback, leading to learning and redesign between sprints.
3. Safe culture. Practice a culture of psychological safety, such that teams have latitude, decision authority, and resources and, above all, feel psychologically safe to innovate, experiment, and fail.

Figure 1. (A) General model of remote spaceflight missions; (B) Framework for Agile and Remote Operations (FAR Ops).



Setting and Participants

The study took place in central Indiana, USA, in partnership with Indiana University, Purdue University, Regenstrief Institute, and local not-for-profit academically affiliated health systems. Participants were consenting cognitively normal adults aged 60 years or older, English-speaking, and using prescribed anticholinergic medications; cognitive normality was determined using the six-item cognitive screener [43]. The first cohort of participants, including all described here, were receiving primary care at Indiana University Health, a statewide academic health system. Individuals were excluded if they were diagnosed with, or using medications to treat, dementia or serious mental illness; were living in an extended care facility or not managing their own medications; were unable to use an app due to sensory disability; or screened positive for cognitive dysfunction or terminal illness. All had consented to the Brain Safe trial and provided a waiver of HIPAA authorization.

Procedure

We applied the Framework for Agile and Remote Operations (FAR Ops) model to achieve technology delivery, support, and return functions in an Agile manner. A team of three study personnel formed the core Project C team: a part-time postdoctoral fellow, with a PhD in industrial and systems engineering; a part-time user experience designer and developer, with an MS in human-computer interaction; and a full-time research specialist, with a BS in neuroscience. The study principal investigator (PI) provided the team with minimal critical specifications to deliver and assess the technology. The team was expected to communicate with the PI, solicit feedback from the broader research team, and develop their own timeline, budget, and procedures. They were explicitly encouraged to be innovative and embrace failure as an opportunity to learn, but to fail early by conducting rapid cycles of testing and redesign, as described in other work on mHealth [44-46].

The team met formally once a week, maintaining informal communications on an as-needed basis, to brainstorm ideas, develop a weekly plan, and monitor the progress of solution development. They worked autonomously and communicated progress to receive feedback and support, not direction, from peers and leaders. The team determined that the first minimum viable product would be a *care package* containing all the

necessary resources and information for participants to receive and use the intervention, along with a word search activity book to show goodwill and increase engagement; this token item was approved by the IRB. In the Results section, we further describe various iterations of the care package product, its delivery, and how users were supported postdelivery.

After the first minimum viable product was ready, the team followed a 2-week sprint schedule. For 1 week they remotely delivered a new batch of packages, and in the subsequent week the team collected feedback, analyzed the results, and redesigned the approach for the next week's deliveries.

To assess technology usability, the team (1) used the Simplified System Usability Scale (SUS) for Cognitively Impaired and Older Adults [47] to obtain a high-level summary usability score and (2) developed a novel usability questionnaire to obtain feedback on specific usability issues with the apps. The novel questionnaire asked for participants' general impressions of the app (eg, if they had trouble reading the text and if they were happy the app was available) as well as their impressions regarding each step of app use (opening the app, logging in, adding and removing medications, etc). Participants were also given the opportunity to express their opinions in their own words in free-text boxes. Participants were given access to an online version of the questionnaire and/or a paper version to be mailed back to the study team. The questionnaires were one way the team was able to return information from remote participants. Another method to return nonstudy data from users, deployed after the first 6 months and, therefore, not reported here, was a user and activity dashboard, with functionality to create or edit accounts remotely, monitor every log-in and in-app function use, and generate custom usage reports.

Analysis

We evaluated the performance of the project on the following criteria, as derived from Agile project management metrics:

1. Time to develop and deliver the first package (ie, minimum viable product); our goal was ≤6 weeks.
2. Adherence to the 2-week sprint cycle; our goal was 100%.
3. Percentage of returning usability questionnaires; no goal was set.

This approach is not textbook Agile. Instead, we have adopted a more general Agile-inspired mindset emphasizing quick progress, autonomous work, and a safe culture within the team. We decided against spending a considerable amount of time implementing more Agile metrics in favor of a quicker response. The three metrics listed above were chosen due to their simplicity and their effectiveness in demonstrating whether our approach was successful in terms of time to respond to the challenge, maintenance of an Agile Mindset, and collection of feedback.

We also analyzed and report feedback obtained from participants and team personnel to evaluate the feasibility of remote operations and the Agile approach. An in-depth analysis of collected usability results was beyond the scope of this paper.

Results

Overview

The 49 participants who were enrolled in the Brain Safe study but had not yet been delivered a study app, were the target of Project C. The demographics of this target group are included in Table 1. No participants had significant cognitive impairments (eg, dementia).

Three main types of care packages were developed and delivered to participants; Types A, B, and C are summarized in Table 2 with the cost of each. Table 3 details the contents and methods used for each package type. Package types, and iterations within type (ie, A_0 vs A_1), were developed between sprints as new needs arose. For example, as some participants began receiving email instructions to download the app to their own smartphone device, the project team learned of the need to provide additional remote installation support.

Table 1. Project C (COVID) target group demographics.

Demographic	Number of participants (N=49), n (%)
Sex	
Male	15 (31)
Female	34 (69)
Age (years)	
60-64	29 (59)
65-69	8 (16)
70-74	8 (16)
75-79	2 (4)
80-84	1 (2)
85+	1 (2)
Race	
White or Caucasian	33 (67)
Black or African American	13 (27)
More than one race	3 (6)
Education level	
Some high school	3 (6)
High school graduate or General Educational Development	8 (16)
Some college	14 (29)
College degree	14 (29)
Master's or other advanced degree	9 (18)
Other or no data	1 (2)

Table 2. Care package types.

Type	Description	Participant needs addressed	Purpose of the package	Cost per package (US $)
A_0	Full package (version 1)	For participants without their own smartphone or preferring to use a study phone	To fulfil the technical (ie, software, hardware, and internet) needs of participants, some of whom were using a mobile app or modern smartphone for the first time	$34.70
A_1	Full package (version 2)	For participants who needed fast and direct delivery	Same as A_0 and to improve the feasibility and lower delivery costs	$17.70
B	Interest-renewal package	For enrolled participants who were difficult to reach in other ways (ie, phone and email)	To re-engage participants who may have lost contact or interest during the pandemic but could not be seen in person	$2.00
C	Remote technical support package	For participants using their own smartphone device and experiencing technical difficulties installing or using the app	To support or troubleshoot self-installation and use of the app	$2.55

Table 3. Care package contents and methods.

Function served and package contents and method	Package type			
	A_0	A_1	B	C
Delivery				
Postal delivery: delivered via United States Postal Service	✔[a]			
Hand delivery: driven to participant's home and left on doorstep by research team member		✔	✔	✔
Study phone: Samsung Galaxy S9 smartphone with charger and protective case	✔	✔		
Getting started guide: documentation with study ID, research and support team contact information, and instructions on first-time phone and app use	✔	✔		
Support				
Technical support contact: IT support phone number added into study phone contacts	✔	✔		
Help documentation: link to online help documentation for the app loaded onto study phone; if requested, printed copy of help documentation was included	✔	✔		
Video walk-through: link to tutorial video with app usage guidance loaded onto study phone; also accessible within the app's help section		✔		
Study phone screen sharing software: Samsung Knox technology management installed on study phones for remote screen sharing and control, with permission	✔	✔		
Other screen sharing software: Zoom or TeamViewer used for remote screen sharing and control, with permission, for participants not using study phones				✔
Return				
Paper System Usability Scale (SUS): hard copy of the SUS	✔	✔		✔
Digital SUS: link to online version of the SUS installed on study phones	✔	✔		
Paper usability questionnaire: hard copy of in-depth usability questionnaire	✔	✔		✔
Digital usability questionnaire: link to online version of in-depth usability questionnaire installed on study phones	✔	✔		
Return envelope: prestamped, preaddressed return envelope to return completed paper questionnaires	✔	✔		✔
Other (ie, goodwill and engagement)				
Token item: word search puzzle book and pencil	✔	✔	✔	✔
Appreciation letter: letter of thanks and explanation from research team to participants		✔	✔	✔

[a]Check marks indicate contents were included in package type.

Delivery

Overview

In the beginning of the pandemic, we faced the challenge of providing mobile apps to older adults who (1) were not tech savvy enough to download and learn to use the app on their own and/or (2) did not possess the technology required to use the app. To address these challenges, two specific conditions related to intervention delivery needed to be met:

1. Smartphones needed to be delivered to those who did not have their own.
2. Participants needed to begin using the study app and smartphone.

Smartphone Delivery

Our need to deliver a physical item (ie, a smartphone) to participants who needed one was the primary reason our team created individual care packages for participants. We initially shipped the packages to participants using the United States Postal Service (USPS). We predicted this would eliminate contact between participants and study personnel, in compliance with institutional pandemic restrictions.

Shipping packages with USPS proved efficient in reducing the time required from Project C team members, but left many delivery aspects outside the control of the study team or inconvenient for participants. Due to the contents of the packages being of high value, the mail recipient's signature was required by USPS. This not only meant that participants needed to be present at the time of delivery, but it also required face-to-face contact between the participant and the postal worker. Each time a participant could not sign for delivery added at least one day's delay. Shipping the packages also affected the cost per package and the time between package creation and delivery.

In the next iteration of the package, the study team decided to *hand-deliver care packages* to participants, thus reducing sign-off failures, costs, and delays. This also led to the team priming participants by confirming their availability to receive the package and their correct address. Project C staff took appropriate disinfecting, distancing, and personal protective

equipment precautions when preparing and hand-delivering packages.

Initiating and Sustaining App Use

Prepandemic protocols had study personnel demonstrate how to use the app to participants during an in-person meeting. To familiarize participants with the operation of the smartphone and the steps required to initiate their use of the app, a *getting started guide* was included in the package. Participants were contacted using their preferred method (ie, email, phone call, or text, as specified during their baseline meeting with the study team).

In addition to package components to facilitate use of the app, we included a *token item* (ie, a word search puzzle book) and *letter of appreciation* from the team to show goodwill in difficult times and add something fun to the package. Due to delays between initial recruitment and app delivery as well as stress that was likely being experienced by many participants at the time, we believed these would promote participant engagement with the team and the study.

We experienced difficulties contacting a subset of participants who had agreed to the study in 2019 but had not yet received the app. In an attempt to reach and re-engage those participants, smaller packages containing only the token item and appreciation letter were sent to 2 such participants. This was discontinued in later sprints in favor of more cost- and time-effective methods of reaching out to nonresponsive participants.

Support

We understood that many participants would require some form of technical support to effectively use the app. We brainstormed many ideas for remote support and over time attempted and refined several. First, we included *contact information for study personnel* (ie, dedicated technical support email and phone number) in each package. We provided in-app one-tap links and a web link to *digital copies of help documentation* demonstrating the features of the app and how to access each app function. Paper help documents were provided as needed.

In later iterations of the package, we created a *video walk-through and tutorial* of the app with voice-over instructions. We determined that seeing the app used on a screen would help some participants learn how to use the app without needing synchronous interactions with study personnel, thus allowing them to use the app at their own pace.

We anticipated that while phone interactions and video tutorials were adequate to address the needs of some participants, there would be others who would benefit from more detailed, personalized support. We implemented *screen sharing capabilities* to offer such support. Study-provided phones were equipped with the option to view and control the phone screen remotely using Knox Manage Remote Support, which was conducted with explicit permission from the participant when needed. This helped when participants could not effectively explain what was on their screen via a phone call and, thus, was viewed as an “invaluable” feature by study personnel.

Additionally, some participants used their own smartphones and were able to download and install the app following email instructions, but required additional remote support. Initially we attempted to use Zoom, a software platform for web-based meetings that allows for screen sharing. We found that installing, registering for, accessing, and using Zoom was too cumbersome for first-time users or caused problems due to system requirements. The team thereafter tested the use of TeamViewer, a similar platform that permits screen sharing by having a participant read a unique 10-digit “Partner ID” to a study team member, leading to on-screen commands to set up and initiate screen sharing. Although TeamViewer required a paid subscription to control the partner’s device, the screen sharing function significantly reduced setup time from over one hour (ie, the time it took for the one attempt to use Zoom) to an average of approximately 10 minutes.

Most privacy concerns about screen sharing were mitigated by openly communicating and building rapport with participants. The screen sharing tools were explained to each participant as well as each step involved in allowing the team member to view their screen. Participants were given opportunities to discuss anything about which they were unsure and it was made clear they could opt out of screen sharing and have technical support delivered solely over the phone instead.

Return

The simplest viable method was to provide participants with a one-tap link to *digital usability assessments* installed on the study phone. Online assessments ensured uniform delivery and collection of responses in one location. However, we hypothesized that some participants would not want to complete questionnaires on a phone. Moreover, for those not using a study phone, we initially considered but rejected options such as mailing a printed Quick Response (QR) code to the assessment link or emailing the link. Therefore, we added *paper usability assessments* to care packages along with a stamped, self-addressed *return envelope*. We theorized that older adult participants would vary in comfort and ability to complete online questionnaires, but all would be able to return mailed questionnaires.

To increase the return of assessments, the team contacted participants by phone and email, when available. In total, 4 participants remained unreachable after multiple attempts.

Evaluation

The first of four packages were sent to participants on April 16, 2020, approximately one month from Project C initiation and within the 6-weeks-or-less target. Subsequent deliveries were maintained on the 2-week schedule through June 2020 per our goal; after a delay, package delivery resumed on the 2-week schedule in late July 2020 (see Table 4). The delay was judged as probably unavoidable but constituted a failure to achieve the target pace.

Table 4. Care package delivery timeline.

Delivery date	Care package type	Number of packages delivered in batch	Within approximately 2 weeks of prior batch?
April 16, 2020	A_0	4	N/A[a]
April 30, 2020	A_1	2	Yes
April 30, 2020	B	2	Yes
May 14, 2020	C	4	Yes
June 4, 2020	A_1	2	Yes
July 23, 2020	A_1	1	No
August 5, 2020	A_1	6	Yes

[a]N/A: not applicable; this was the first delivery date.

During the 6-month period, 19 participants were invited to complete usability assessments in their care packages. Of these, 10 (53%) returned the short, 10-item SUS. Of those 10 participants, 6 (60%) used the link on the study phone, 2 (20%) were emailed the link after providing an email address, and 2 (20%) completed the questionnaire by phone after receiving a follow-up phone call. None successfully returned a mailed SUS, though we do not know how many attempted without success, for example, if their response was lost in the mail due to campus reorganization of mail delivery services. Out of 19 participants, 8 (42%) returned the longer, in-depth questionnaire to the study team. Of those, 5 (63%) used the in-phone link, 2 (25%) completed by phone, and 1 (13%) both returned a paper version and completed the questionnaire online.

Providing email links or having participants respond to assessment questionnaires over the phone are examples of our iterative process enabling our team to respond and adjust our approach based on the unsolicited feedback of our participants. We did not initially have email addresses for our participants and we anticipated that the use of USPS and paper questionnaires would be utilized more frequently than they were. When participants requested that they be allowed to fill out the questionnaire through email, and provided the team an email address, or indicated they had time to discuss the questionnaire and provide answers over the phone, we were able to adjust to meet their needs. While these changes did not constitute a new care package iteration, they are still agile adjustments that our team was able to make as we gained more information about the preferences of our participants.

Discussion

Principal Findings

During the COVID-19 pandemic, some clinical trials were forced to shut down [48], while others found ways to adapt [49]. To adapt technology-based research, we created the FAR Ops framework, based on spaceflight operations and Agile processes. Guided by the framework, we iteratively developed and tested solutions, which enabled us to deliver a study intervention and collect usability feedback from older adults, remotely. We succeeded in designing and deploying a minimum viable product in less than 6 weeks, generally maintained a 2-week sprint cycle, and achieved a 40% to 50% return rate for usability assessment instruments.

Our work demonstrates the feasibility of "going remote" to deliver and assess technology interventions to a diverse population during a pandemic and likely in future, less dire circumstances. Researchers and practitioners can replicate our approach as an alternative to face-to-face interactions. Doing so not only permits work to continue during circumstances such as communicable disease outbreaks and natural disasters, but can also save costs, accommodate people with disability or other reasons for staying at home (eg, caregiving responsibilities), reduce travel time and reliance on transportation for participants and staff, allow for testing in natural settings, and overall increase the diversity of participants [11,12,18,50].

The FAR Ops conceptual model provides a framework for such endeavors. Its Agile project management approach can be replicated to ensure goals are met on time, progress is made at a rapid pace, and innovation is embedded in the process by continually refining the solution based on emerging needs and feedback from testing [40]. We demonstrated how simple Agile concepts, such as explicitly setting expectations for a rapid first "good enough" iteration or minimum viable product and short sprint cycles, resulted in rapid, continuous progress.

Our Agile approach also helped the team iteratively respond to new information, which led to adaptations in package design to meet needs, solve problems, and improve effectiveness and efficiency. This is not possible in a traditional or *waterfall* approach to project management, wherein solution development is front-loaded and prolonged in an effort to come up with the ideal solution, leaving little time, opportunity, and resources available to redesign and retest the solution if it turns out to not be as ideal as assumed. In fact, it has been documented that technologies developed without integrating iterative testing and redesign opportunities result in products that do not effectively respond to user needs [51-53]. Compared to waterfall approaches, Agile allows feedback from customers—in our case, participants—to dictate changes, rather than assuming the design team can predict customer experiences [54]. However, because an Agile approach promotes quickly creating the minimum viable product, this can result in early solutions having fewer features. This was true in our case, as participants in earlier iterations did not receive tutorial videos or screen sharing, leading to challenges. At the same time, these challenges inspired these subsequent features to be added, reinforcing the importance of the Agile approach for innovation and

determining the contents of the solution based on actual experimentation [55].

We adopted an Agile Mindset, meaning we went beyond specific Agile techniques and even downplayed the need to adopt specific Agile techniques, such as measuring velocity and throughput, diagramming, or having formal scrum teams [56,57]. We emphasized instead the Agile Mindset principle of establishing a psychologically safe culture within which the team was empowered to work autonomously, make quick progress, prefer "good enough" to "perfect," and learn from failures. However, we acknowledge by some standards, the project's specific methods were discordant with textbook Agile approaches.

Much like user-centered design, our approach was iterative, catered to diverse participant needs, and made changes to designs based on evidence provided by users [51-53]. We assert that user-centered design and the Agile approach are complementary—the Agile approach can help operationalize user-centered design best practices, while user-centered design compels solutions that fit the end user. The concept of the Agile approach and user-centered design compatibility is detailed in Holden and Boustani [58] and Holden et al [59].

A significant challenge overcome by our study team was the lack of existing methodologies and guidelines available for remote usability testing with older adults. Recently, McLaughlin et al [8] outlined various tools available for different kinds of remote usability testing of medical devices and recommended that study personnel be available for phone support during tests with older adults. Beyond this advice, there is no clear guidance from human-computer interaction, usability, mHealth, or other literature on how to operationalize a fully remote study protocol for older adult participants. Based on the results of our work, we present additional recommendations for remote operations with diverse populations in Table 5.

Table 5. Recommendations for remote usability studies with diverse populations.

Recommendation	Details
Screen sharing and screen control	• Valuable for technical support • Use alongside phone-based support • Allows research personnel to more clearly understand participant challenges • Use screen sharing platforms that minimize input required from participant (eg, Knox Manage Remote Support and TeamViewer)
Phone-based interactions	• Some participants may prefer to be contacted by telephone • Allows for participants to talk through impressions and responses to usability evaluations • Can enable researchers to gather more detailed responses to open-ended questions • Determine if participant would prefer to answer questionnaires over the phone
Email	• Some participants may be very familiar with email interactions while others are not • Determine if participant has a valid email and would prefer to be emailed a link to a usability questionnaire, over other methods

About 50% of participants returned their usability assessments, fewer than desired and, thus, limiting the success of the project. This may be due to reduced levels of engagement overall, especially during unusual circumstances (eg, the COVID-19 pandemic), or as a result of remote and often asynchronous communication. Prior work recommends one-on-one, synchronous conversation as the most effective method to elicit information from older adults [27,28] and may explain why some participants preferred to complete assessments by phone, even though this option was not initially offered. Synchronous interactions may have been preferred by participants to provide lengthier or supplemental feedback to the team. Indeed, phone responses were, on average, longer than those through self-administered modes. We further hypothesize that some participants may have had difficulty typing their responses or using online questionnaires, which contributed to nonresponse or preference to complete assessments by phone, though this was not communicated to the study team by any participants. This may have been especially so for the longer questionnaire, although response rates for the shorter structured 10-item SUS (53%) and the longer questionnaire with free-text responses (42%) were not exceptionally different. It is important to note that while this flexibility in data collection methodology enabled us to collect more responses from participants, it also introduced variability in the level of detail of collected responses. This could challenge the fidelity of studies that heavily rely on consistent data collection techniques.

Maintaining participant engagement was difficult for our team and may indicate that it would be valuable if, in the future, we were more closely able to replicate face-to-face interactions when assessing intervention usability. We are currently exploring adapting existing synchronous remote usability methods for older adults.

The ages of the 19 participants who received Type A and C packages were, overall, representative of the larger target group: 12 in their 60s, 5 in their 70s, and 1 participant in their 80s. Younger participants were more likely than older participants to return usability assessments to the study team; of those older participants who did return assessments, they were more likely to prefer phone interactions over electronic responses. The two SUS assessments performed over the phone were collected from participants in their 70s, while those in their 60s returned the surveys electronically. Only 1 individual in their 70s completed the longer usability questionnaire, while the rest were returned by younger participants. Although any conclusions are speculative given the small sample size, it may be easier to "go remote" with younger older adults.

A limitation of this study is that it was not designed to compare our approach with others, prohibiting conclusions about its superiority. We also did not design our study to compare various package iterations to each other. During the project, decisions on design iterations were based on relatively small samples of participant feedback and impressions of study personnel. We report only 6 months of the project. Our sample was relatively diverse, but overrepresented younger, White, and female individuals. We believe that limitations with our sample are reasonable due to the nature of the COVID-19 pandemic. We did not formally assess the team's fidelity to Agile principles or use certain Agile formalisms, such as velocity or control charting, to track progress.

Conclusions

Combining a spaceflight model with an Agile approach allowed our team to adopt remote operations for our mHealth trial, in response to interruptions from the COVID-19 pandemic. Our approach can be useful for other research or practical projects under similar circumstances or to improve efficiency, cost, effectiveness, and participant diversity in general. Indeed, although the COVID-19 pandemic was the impetus for "going remote," we believe further iterating remote operations for older adults will result in these approaches being preferable to traditional ones for many reasons.

In addition to offering a replicable approach, this paper tells the often-untold story of practical challenges faced by mHealth projects and practical strategies used, as another article in this journal [45] compels us to do:

> *Telling and learning from the typically untold stories will result in more efficient, effective, and sustainable mHealth design efforts...We call on our fellow researchers, designers, and UCD [user-centered design] experts to document and share their own challenges and strategies toward improving the implementation of UCD.*

Acknowledgments

We thank members of the Brain Safety Lab for their input. The authors of this paper were supported by grant R01AG056926 from the National Institute on Aging of the National Institutes of Health (NIH). The content is solely the responsibility of the authors and does not necessarily represent the official views of NIH.

Authors' Contributions

JRH and RJH conceived this paper. All authors wrote and edited the paper. All authors approved the final version.

Conflicts of Interest

None declared.

References

1. Messenger JC. Telework in the 21st Century: An Evolutionary Perspective. Northampton, MA: Edward Elgar Publishing, Inc; Nov 29, 2019.
2. HHS issues new report highlighting dramatic trends in Medicare beneficiary telehealth utilization amid COVID-19. US Department of Health & Human Services. 2020 Jul 28. URL: https://www.hhs.gov/about/news/2020/07/28/hhs-issues-new-report-highlighting-dramatic-trends-in-medicare-beneficiary-telehealth-utilization-amid-covid-19.html [accessed 2021-02-25]
3. Loeb AE, Rao SS, Ficke JR, Morris CD, Riley LH, Levin AS. Departmental experience and lessons learned with accelerated introduction of telemedicine during the COVID-19 crisis. J Am Acad Orthop Surg 2020;28(11):e469-e476. [doi: 10.5435/jaaos-d-20-00380]
4. Andrienko O. Semrush Blog. 2020 Apr 20. URL: https://www.semrush.com/blog/ecommerce-covid-19/ [accessed 2021-02-25]
5. Lai B, Chiu C, Pounds E, Tracy T, Mehta T, Young H, et al. COVID-19 modifications for remote teleassessment and teletraining of a complementary alternative medicine intervention for people with multiple sclerosis: Protocol for a randomized controlled trial. JMIR Res Protoc 2020 Jul 03;9(7):e18415 [FREE Full text] [doi: 10.2196/18415] [Medline: 32540838]
6. DeMatthews D, Knight D, Reyes P, Benedict A, Callahan R. From the field: Education research during a pandemic. Educ Res 2020 Jul 07;49(6):398-402. [doi: 10.3102/0013189x20938761]
7. Omary MB, Eswaraka JR, Kimball SD, Moghe PV, Panettieri RA, Scotto KW. The COVID-19 pandemic and research shutdown: Staying safe and productive. J Clin Invest 2020 Jun 01;130(6):2745-2748 [FREE Full text] [doi: 10.1172/JCI138646] [Medline: 32243259]
8. McLaughlin AC, DeLucia PR, Drews FA, Vaughn-Cooke M, Kumar A, Nesbitt RR, et al. Evaluating medical devices remotely: Current methods and potential innovations. Hum Factors 2020 Nov;62(7):1041-1060 [FREE Full text] [doi: 10.1177/0018720820953644] [Medline: 32960093]
9. Hartson HR, Castillo JC. Remote evaluation for post-deployment usability improvement. In: Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '98). New York, NY: Association for Computing Machinery; 1998 May Presented at: Working Conference on Advanced Visual Interfaces (AVI '98); May 24-27, 1998; L'Aquila, Italy p. 22-29

URL: https://static.aminer.org/pdf/PDF/000/061/037/remote_evaluation_for_post_deployment_usability_improvement.pdf [doi: 10.1145/948496.948499]
10. Castillo JC, Hartson HR, Hix DH. Remote usability evaluation: Can users report their own critical incidents? In: Proceedings of the Conference on Human Factors and Computing Systems (CHI '98). New York, NY: Association for Computing Machinery; 1998 Apr Presented at: Conference on Human Factors and Computing Systems (CHI '98); April 18-23, 1998; Los Angeles, CA p. 253-254. [doi: 10.1145/286498.286736]
11. Andreasen M, Nielsen H, Schrøder S, Stage J. What happened to remote usability testing? An empirical study of three methods. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07). New York, NY: Association for Computing Machinery; 2007 Apr Presented at: SIGCHI Conference on Human Factors in Computing Systems (CHI '07); April 28-May 3, 2007; San Jose, CA p. 1405-1414. [doi: 10.1145/1240624.1240838]
12. Hammontree M, Weiler P, Nayak N. Remote usability testing. Interactions 1994 Jul;1(3):21-25. [doi: 10.1145/182966.182969]
13. Tullis T, Fleischman S, Mcnulty M, Cianchette C, Bergel M. An empirical comparison of lab and remote usability testing of web sites. In: Proceedings of the 11th Usability Professionals Association Conference. 2002 Jul Presented at: 11th Usability Professionals Association Conference; July 8-12, 2002; Orlando, FL URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.457.3080&rep=rep1&type=pdf
14. Bernheim Brush AJ, Ames M, Davis J. A comparison of synchronous remote and local usability studies for an expert interface. In: Proceedings of the Conference on Human Factors in Computing Systems (CHI04). New York, NY: Association for Computing Machinery; 2004 Apr Presented at: Conference on Human Factors in Computing Systems (CHI04); April 24-29, 2004; Vienna, Austria p. 1179-1182. [doi: 10.1145/985921.986018]
15. Diamantidis CJ, Ginsberg JS, Yoffe M, Lucas L, Prakash D, Aggarwal S, et al. Remote usability testing and satisfaction with a mobile health medication inquiry system in CKD. Clin J Am Soc Nephrol 2015 Aug 07;10(8):1364-1370 [FREE Full text] [doi: 10.2215/CJN.12591214] [Medline: 26220816]
16. Bruun A, Gull P, Hofmeister L, Stage J. Let your users do the testing: A comparison of three remote asynchronous usability testing methods. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09). New York, NY: Association for Computing Machinery; 2009 Apr Presented at: SIGCHI Conference on Human Factors in Computing Systems (CHI '09); April 4-9, 2009; Boston MA p. 1619-1628. [doi: 10.1145/1518701.1518948]
17. Fidas C, Katsanos C, Papachristos E, Tselios N, Avouris N. Remote usability evaluation methods and tools: A survey. In: Proceedings of the 11th Panhellenic Conference on Informatics (PCI 2007). 2007 Presented at: 11th Panhellenic Conference on Informatics (PCI 2007); May 18-20, 2007; Patras, Greece p. 151-162 URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.545.7719&rep=rep1&type=pdf
18. Petrie H, Hamilton F, King N, Pavan P. Remote usability evaluations with disabled people. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06). New York, NY: Association for Computing Machinery; 2006 Apr Presented at: SIGCHI Conference on Human Factors in Computing Systems (CHI '06); April 22-27, 2006; Montréal, QC p. 1133-1141. [doi: 10.1145/1124772.1124942]
19. Anderson M, Perrin A. Tech adoption climbs among older adults. Pew Research Center. Washington, DC: Pew Research Center; 2017 May 17. URL: https://www.pewresearch.org/internet/2017/05/17/tech-adoption-climbs-among-older-adults/ [accessed 2021-02-25]
20. Wildenbos GA, Peute L, Jaspers M. Aging barriers influencing mobile health usability for older adults: A literature based framework (MOLD-US). Int J Med Inform 2018 Jun;114:66-75. [doi: 10.1016/j.ijmedinf.2018.03.012] [Medline: 29673606]
21. Ramsetty A, Adams C. Impact of the digital divide in the age of COVID-19. J Am Med Inform Assoc 2020 Jul 01;27(7):1147-1148 [FREE Full text] [doi: 10.1093/jamia/ocaa078] [Medline: 32343813]
22. Internet/broadband fact sheet. Pew Research Center. Washington, DC: Pew Research Center; 2019 Jun 12. URL: https://www.pewresearch.org/internet/fact-sheet/internet-broadband/ [accessed 2021-02-25]
23. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: Internet use patterns, eHealth literacy, and attitudes toward computer/internet use. J Med Internet Res 2013 May 02;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
24. Mitchell U, Chebli P, Ruggiero L, Muramatsu N. The digital divide in health-related technology use: The significance of race/ethnicity. Gerontologist 2019 Jan 09;59(1):6-14. [doi: 10.1093/geront/gny138] [Medline: 30452660]
25. Smith A. Older adults and technology use. Pew Research Center. Washington, DC: Pew Research Center; 2014 Apr 03. URL: https://www.pewresearch.org/internet/2014/04/03/older-adults-and-technology-use/ [accessed 2021-02-25]
26. Levy H, Janke AT, Langa KM. Health literacy and the digital divide among older Americans. J Gen Intern Med 2015 Mar;30(3):284-289 [FREE Full text] [doi: 10.1007/s11606-014-3069-5] [Medline: 25387437]
27. Dickinson A, Arnott J, Prior S. Methods for human – Computer interaction research with older people. Behav Inf Technol 2007 Jul;26(4):343-352. [doi: 10.1080/01449290601176948]
28. Franz R, Munteanu C, Barbosa NB, Baecker R. Time to retire old methodologies? Reflecting on conducting usability evaluations with older adults. In: Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI '15). New York, NY: Association for Computing Machinery; 2015 Aug Presented at: 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI '15); August 24-27, 2015; Copenhagen, Denmark p. 912-915. [doi: 10.1145/2786567.2794303]

29. Silva P, Nunes F. 3 x 7 usability testing guidelines for older adults. In: Proceedings of the 3rd Mexican Workshop on Human Computer Interaction (MexIHC '10). San Luis Potosí, Mexico: Universidad Politécnica de San Luis Potosí; 2010 Nov Presented at: 3rd Mexican Workshop on Human Computer Interaction (MexIHC '10); November 8-10, 2010; San Luis Potosí, Mexico p. 1-8 URL: http://mural.maynoothuniversity.ie/6030/1/PS-Usability-Testing.pdf
30. Abebe E, Campbell NL, Clark DO, Tu W, Hill JR, Harrington AB, et al. Reducing anticholinergic medication exposure among older adults using consumer technology: Protocol for a randomized clinical trial. Res Soc Adm Pharm 2020 Oct:1 (forthcoming). [doi: 10.1016/j.sapharm.2020.10.010]
31. Holden RJ, Campbell NL, Abebe E, Clark DO, Ferguson D, Bodke K, Brain Health Patient Safety Laboratory. Usability and feasibility of consumer-facing technology to reduce unsafe medication use by older adults. Res Social Adm Pharm 2020 Jan;16(1):54-61 [FREE Full text] [doi: 10.1016/j.sapharm.2019.02.011] [Medline: 30853507]
32. Howell E. A brief history of Mars missions. Space. 2019. URL: https://www.space.com/13558-historic-mars-missions.html [accessed 2021-02-25]
33. Humans in space overview. National Aeronautics and Space Administration (NASA). 2018 Jul 25. URL: https://www.nasa.gov/topics/humans-in-space/overview [accessed 2021-02-25]
34. Missions. Jet Propulsion Laboratory, NASA. URL: https://www.jpl.nasa.gov/missions/?type=current [accessed 2021-02-25]
35. Mishkin A. Sojourner: An Insider's View of the Mars Pathfinder Mission. New York, NY: The Berkley Publishing Group; 2003.
36. Woolford B, Sipes W, Fiedler E. Human space flight. In: Salvendy G, editor. Handbook of Human Factors and Ergonomics. Fourth edition. Hoboken, NJ: John Wiley & Sons, Inc; 2012:910-927.
37. Bell E, Coan D, Oswald D. A discussion on the making of an EVA: What it really takes to walk in space. In: Proceedings of the SpaceOps 2006 Conference. 2006 Presented at: SpaceOps 2006 Conference; June 19-23, 2006; Rome, Italy p. 1-20 URL: https://arc.aiaa.org/doi/pdf/10.2514/6.2006-5570 [doi: 10.2514/6.2006-5570]
38. Heiney A. SpaceX-3 science payloads return to Kennedy. National Aeronautics and Space Administration (NASA). 2014 May 23. URL: https://www.nasa.gov/content/spacex-3-science-payloads-return-to-kennedy [accessed 2021-02-25]
39. Dove R. Knowledge management, response ability, and the agile enterprise. J Knowl Manag 1999 Mar;3(1):18-35. [doi: 10.1108/13673279910259367]
40. Beck K, Beedle M, van Bennekum A, Cockburn A, Cunningham W, Fowler M, et al. Manifesto for agile software development. Agile Manifesto. 2001. URL: https://agilemanifesto.org/ [accessed 2021-02-25]
41. Boustani MA, van der Marck MA, Adams N, Azar JM, Holden RJ, Vollmar HC, et al. Developing the agile implementation playbook for integrating evidence-based health care services into clinical practice. Acad Med 2019;94(4):556-561. [doi: 10.1097/acm.0000000000002497]
42. Holden RJ, Boustani MA. The value of an "agile" mindset in times of crisis. Modern Healthcare. 2020 May 11. URL: https://www.modernhealthcare.com/opinion-editorial/value-agile-mindset-times-crisis [accessed 2021-02-25]
43. Callahan CM, Unverzagt FW, Hui SL, Perkins AJ, Hendrie HC. Six-item screener to identify cognitive impairment among potential subjects for clinical research. Med Care 2002 Sep;40(9):771-781. [doi: 10.1097/00005650-200209000-00007] [Medline: 12218768]
44. Holden RJ, Bodke K, Tambe R, Comer RS, Clark DO, Boustani M. Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care. 2016 Jul 22 Presented at: International Symposium on Human Factors and Ergonomics in Health Care; April 13-16, 2016; San Diego, CA p. 25-27. [doi: 10.1177/2327857916051003]
45. Cornet VP, Toscos T, Bolchini D, Rohani Ghahari R, Ahmed R, Daley C, et al. Untold stories in user-centered design of mobile health: Practical challenges and strategies learned from the design and evaluation of an app for older adults with heart failure. JMIR Mhealth Uhealth 2020 Jul 21;8(7):e17703 [FREE Full text] [doi: 10.2196/17703] [Medline: 32706745]
46. Srinivas P, Cornet V, Holden R. Human factors analysis, design, and evaluation of Engage, a consumer health IT application for geriatric heart failure self-care. Int J Hum Comput Interact 2017;33(4):298-312 [FREE Full text] [doi: 10.1080/10447318.2016.1265784] [Medline: 30429638]
47. Holden RJ. A Simplified System Usability Scale (SUS) for Cognitively Impaired and Older Adults. In: Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care. 2020 Sep 16 Presented at: International Symposium on Human Factors and Ergonomics in Health Care; March 8-11, 2020; Toronto, ON p. 180-182. [doi: 10.1177/2327857920091021]
48. Thornton J. Clinical trials suspended in UK to prioritise COVID-19 studies and free up staff. BMJ 2020 Mar 23;368:m1172. [doi: 10.1136/bmj.m1172] [Medline: 32205354]
49. Verna EC, Serper M, Chu J, Corey K, Fix OK, Hoyt K, et al. Clinical research in hepatology in the COVID-19 pandemic and post-pandemic era: Challenges and the need for innovation. Hepatology 2020 Nov;72(5):1819-1837 [FREE Full text] [doi: 10.1002/hep.31491] [Medline: 32740969]
50. Dray S, Siegel D. Remote possibilities?: International usability testing at a distance. Interactions 2004 Mar;11(2):10-17. [doi: 10.1145/971258.971264]
51. Goldberg L, Lide B, Lowry S, Massett HA, O'Connell T, Preece J, et al. Usability and accessibility in consumer health informatics current trends and future challenges. Am J Prev Med 2011 May;40(5 Suppl 2):S187-S197. [doi: 10.1016/j.amepre.2011.01.009] [Medline: 21521594]

52. Marquard JL, Zayas-Cabán T. Commercial off-the-shelf consumer health informatics interventions: Recommendations for their design, evaluation and redesign. J Am Med Inform Assoc 2012;19(1):137-142 [FREE Full text] [doi: 10.1136/amiajnl-2011-000338] [Medline: 21727206]
53. Zayas-Cabán T, Dixon BE. Considerations for the design of safe and effective consumer health IT applications in the home. Qual Saf Health Care 2010 Oct;19 Suppl 3:i61-i67. [doi: 10.1136/qshc.2010.041897] [Medline: 20959321]
54. VanderLeest S, Buter A. Escape the waterfall: Agile for aerospace. In: Proceedings of the 2009 IEEE/AIAA 28th Digital Avionics Systems Conference. 2009 Presented at: 2009 IEEE/AIAA 28th Digital Avionics Systems Conference; October 23-29, 2009; Orlando, FL p. 6.D.3-1-6.D.3-16. [doi: 10.1109/dasc.2009.5347438]
55. Boustani M, Holden R, Azar J, Solid C. The Agile Network: A Model to Foster Innovation, Implementation, and Diffusion in Healthcare Systems. St Paul, MN: Beaver's Pond Press; 2019.
56. Radigan D. Five agile metrics you won't hate. Atlassian Agile Coach. URL: https://www.atlassian.com/agile/project-management/metrics [accessed 2021-02-25]
57. Chowdhury A. Agile metrics: The 15 that actually matter for success. Plutora. 2020. URL: https://www.plutora.com/blog/agile-metrics [accessed 2021-02-25]
58. Holden RJ, Boustani MA. Design and usability methods: Agile innovation and evaluation of interventions for patients and families. In: Holden RJ, Valdez RS, editors. The Patient Factor: Theories and Methods for Patient Ergonomics. Boca Raton, FL: CRC Press; 2021.
59. Holden RJ, Boustani MA, Azar J. Agile innovation to transform healthcare: Innovating in complex adaptive systems is an everyday process, not a light bulb event. BMJ Innov 2021 Jan 28:1-7 [FREE Full text] [doi: 10.1136/bmjinnov-2020-000574]

Abbreviations

FAR Ops: Framework for Agile and Remote Operations
HIPAA: Health Insurance Portability and Accountability Act
IRB: Institutional Review Board
mHealth: mobile health
NIH: National Institutes of Health
PI: principal investigator
Project C: Project COVID
QR: Quick Response
SUS: System Usability Scale
UCD: user-centered design
USPS: United States Postal Service

Edited by L Buis; submitted 22.12.20; peer-reviewed by C Mayhorn, K Yin, M Mirro; comments to author 06.01.21; revised version received 18.01.21; accepted 17.02.21; published 04.03.21.

Please cite as:
Hill JR, Harrington AB, Adeoye P, Campbell NL, Holden RJ
Going Remote—Demonstration and Evaluation of Remote Technology Delivery and Usability Assessment With Older Adults: Survey Study
JMIR Mhealth Uhealth 2021;9(3):e26702
URL: https://mhealth.jmir.org/2021/3/e26702
doi:10.2196/26702
PMID:33606655

Original Paper

Learning From Others Without Sacrificing Privacy: Simulation Comparing Centralized and Federated Machine Learning on Mobile Health Data

Jessica Chia Liu[1], MA; Jack Goetz[1], PhD; Srijan Sen[2,3], MD, PhD; Ambuj Tewari[1], PhD

[1]Department of Statistics, University of Michigan, Ann Arbor, MI, United States

[2]Molecular and Behavioral Neuroscience Institute, University of Michigan, Ann Arbor, MI, United States

[3]Department of Psychiatry, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Jessica Chia Liu, MA
Department of Statistics
University of Michigan
1085 South University Ave
Ann Arbor, MI, 48109
United States
Phone: 1 7346474820
Email: liujess@umich.edu

Abstract

Background: The use of wearables facilitates data collection at a previously unobtainable scale, enabling the construction of complex predictive models with the potential to improve health. However, the highly personal nature of these data requires strong privacy protection against data breaches and the use of data in a way that users do not intend. One method to protect user privacy while taking advantage of sharing data across users is federated learning, a technique that allows a machine learning model to be trained using data from all users while only storing a user's data on that user's device. By keeping data on users' devices, federated learning protects users' private data from data leaks and breaches on the researcher's central server and provides users with more control over how and when their data are used. However, there are few rigorous studies on the effectiveness of federated learning in the mobile health (mHealth) domain.

Objective: We review federated learning and assess whether it can be useful in the mHealth field, especially for addressing common mHealth challenges such as privacy concerns and user heterogeneity. The aims of this study are to describe federated learning in an mHealth context, apply a simulation of federated learning to an mHealth data set, and compare the performance of federated learning with the performance of other predictive models.

Methods: We applied a simulation of federated learning to predict the affective state of 15 subjects using physiological and motion data collected from a chest-worn device for approximately 36 minutes. We compared the results from this federated model with those from a centralized or server model and with the results from training individual models for each subject.

Results: In a 3-class classification problem using physiological and motion data to predict whether the subject was undertaking a neutral, amusing, or stressful task, the federated model achieved 92.8% accuracy on average, the server model achieved 93.2% accuracy on average, and the individual model achieved 90.2% accuracy on average.

Conclusions: Our findings support the potential for using federated learning in mHealth. The results showed that the federated model performed better than a model trained separately on each individual and nearly as well as the server model. As federated learning offers more privacy than a server model, it may be a valuable option for designing sensitive data collection methods.

(JMIR Mhealth Uhealth 2021;9(3):e23728) doi:10.2196/23728

KEYWORDS

privacy; data protection; machine learning; mobile health; wearable electronic devices

Introduction

Mobile Health Data

The ubiquitous nature of wearables generates considerable potential for data collection and analysis but comes with the issue of protecting user privacy. Some important privacy considerations are protecting patient confidentiality [1], protecting against security breaches [2], and protecting against researchers using user data in a way that the user did not intend [3,4]. Individuals may be more willing to participate in studies and disclose information if these concerns are alleviated [5,6]. Privacy breaches can occur when data servers are compromised but can also occur when data are shared for legitimate purposes by well-intentioned members of the medical community [7]. In this paper, we use privacy to denote the aspects related to protecting the identity, personal information, and use of the data of users.

Mobile health (mHealth) data are often related to the cognitive, behavioral, and affective states of users, making such data highly sensitive. Therefore, we want to ensure that individuals' confidential health information is not leaked to others. Wearables passively record a range of medically relevant data, such as temperature, heart rate, and electrodermal activity (EDA). As people may carry these items with them throughout the day, this allows for high-frequency collection of data from more people, who may have a greater variety of health conditions, than ever before. Such rich data collection opens up the possibility of using increasingly powerful but data hungry machine learning methods in the analysis of these data [8].

In this paper, we apply predictive machine learning models on the publicly available *Wearable Stress and Affect Detection* (WESAD) data set published by Schmidt et al [9]. In particular, we focus on models that can be trained by fitting a function, potentially nonlinear, to the data using gradient descent and variants thereof. Many of these models make few assumptions about the structure of the underlying data-generating process. To improve privacy, we propose leaving each user's data on their personal device and training our models using federated learning. In federated learning, there is no single server that contains all users' information. Instead, model training occurs on each individual's device, and only model parameter updates leave the user's device. This allows for more user privacy by maintaining data only on individual user devices. In addition, as explained later, federated learning is still able to take advantage of some shared information across individuals. Thus, it can alleviate some of the concerns in analyzing mHealth data, such as user heterogeneity and privacy preservation.

Federated Learning

Before further describing federated learning, we establish some common terminology. The goal is to produce a model, often a neural network, trained using data from many individuals. Each individual will have a mobile device, which we call a *user device*—in the networking literature, these are called *clients*—and we will not make a distinction between an individual and their user device. Each user device has its own private data, and the data will never leave the user device. Storing private data on each user device instead of uploading them centrally is the source of improved privacy provided by federated learning. We will also have a single device controlled by the researchers that can communicate with each user device, and we will call the former device the *server.* The server will coordinate the training procedure and store a copy of the model but will not have any private data uploaded onto it. The training process is iterative, as is common for many machine learning models. Each time the model parameters on the server change, we will call that one server training round (or server training iteration).

In Figure 1, we compare how a model is trained on a central server (top) with how a model is trained using federated learning (bottom). In federated learning, one server training round comprises the following three distinct parts:

1. Broadcast: A small number of user devices will be selected at random, and the current server model will be transmitted to that cohort of user devices.
2. Local update: Each of these user devices will perform a small amount of training of the model they received, using the data from the user device.
3. Update aggregation: Each user device in this cohort will then transmit a copy of their (locally updated) model parameters to the server. The server then averages these parameters and replaces the server model parameters with these new averaged parameters.

The server then repeats this process by selecting a new cohort of user devices for each server training round. Thus, the data from every user device contribute to the training of the model. Algorithm 1 (Figure 2) presents the pseudocode, and a full algorithmic description is provided in Multimedia Appendix 1.

Figure 1. Diagram comparing central and federated learning workflows. Color abstractly represents the private information content at different locations, with red, blue, and green colors representing private information for different user devices. When training on a central server, user data are uploaded onto the server once. In federated learning, model parameters are updated on the user device, producing updates that contain less private information than the data themselves. The updates from many users are then aggregated, further mixing the contribution from each individual user.

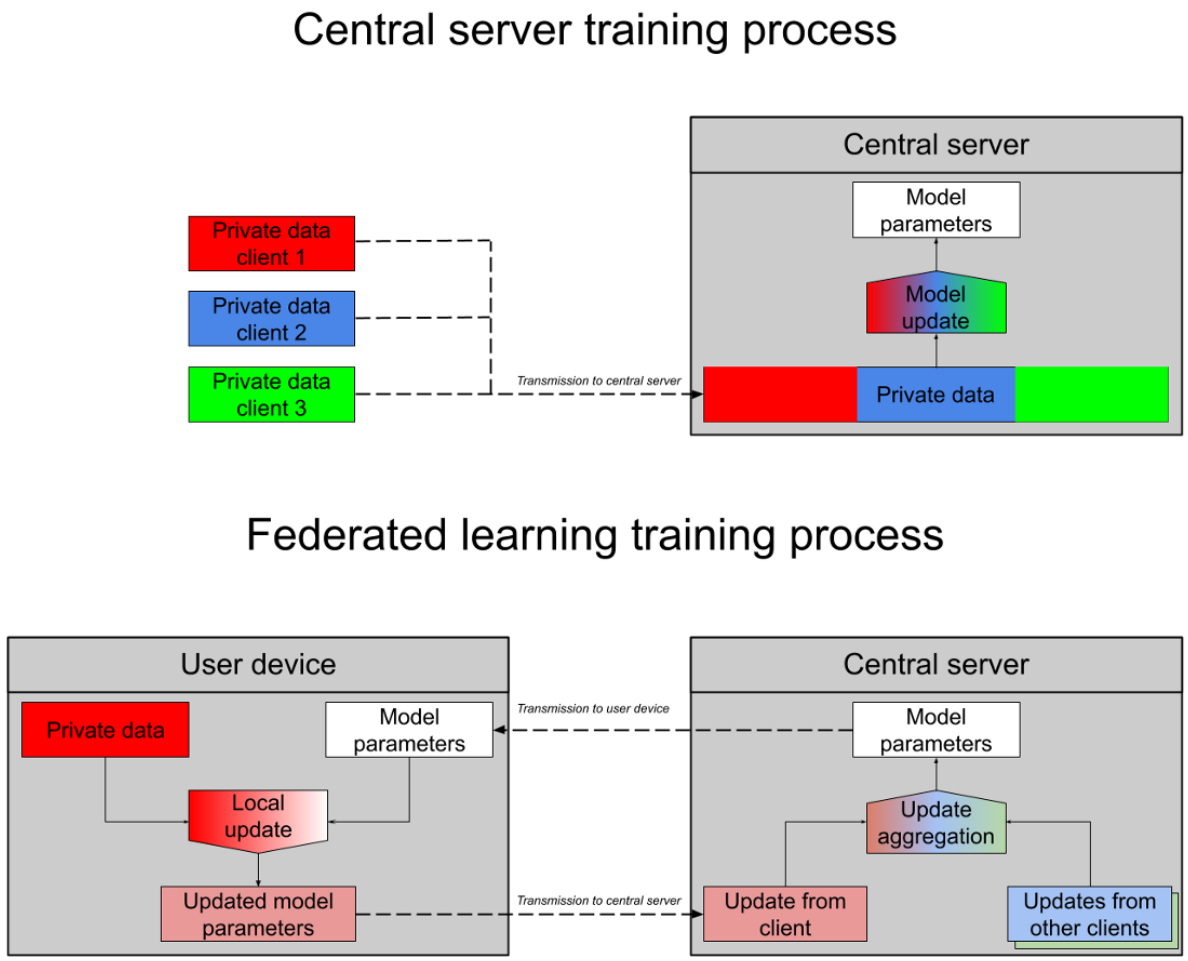


Figure 2. Pseudocode for federated learning algorithm 1.

Algorithm 1: Federated learning pseudocode

Input: Central server S, number of clients K, cohort size k, number of server training rounds T, initial model parameters θ_0.

Output: Final model parameters θ_T

for t *from* 1 *to* T **do**

  Select k out of K at random to form cohort $C_t \subset \{1, .., K\}$.

  for i *in* C_t **do**

    Transmit current model parameters θ_{t-1} from S to client i.

    Client i updates the model parameters θ_{t-1} using their data to produce $\theta_t^{(i)}$.

    Transmit $\theta_t^{(i)}$ back to the server S.

  end

  Server S averages the $\theta_t^{(i)}$ from all clients in cohort, producing θ_t.

end

Output final model parameters θ_T.

The model trains using updates from the data on all user devices; however, individual, private, data remain only with each user device. This protects the privacy of users in several ways. Under federated learning, private data are significantly less vulnerable to the central server being compromised. The server never contains any raw private data and generally contains only the current model parameters. This protects users from being subject to adverse consequences of failings in the server's security. Another benefit for users under federated learning is private data cannot be reused for other purposes at a later date. It cannot be shared or released either accidentally or in good faith but with inadequate anonymization measures. Finally, federated learning gives users the freedom to withdraw access to their data—both past and future—at any time without relying on guarantees from the researchers that any previously collected data will be deleted.

Federated learning has been successfully applied in complex real-world applications, most notably in training the next-word prediction model for Google's Keyboard [10]. We refer interested readers to excellent surveys of federated learning [11-13] and federated learning software and data sets [13].

Previous Work

There has been substantial research on privacy in mHealth and mood prediction using machine learning. Protecting the data of individual users stored in mHealth apps is important because the apps contain personal information about the user, information used to make treatment decisions, and information that could be misused for financial gain, among other concerns [4]. Users' willingness to disclose personal information in mobile data varies according to their demographics and personal characteristics [14-16]. Concerns about privacy vary among different age groups; for example, some older people are less willing to use mHealth services [6,14] and are more likely to cite privacy concerns [6]. As such, it is important to consider the target population when considering the amount of privacy guarantee needed. Increased privacy concerns about health information technologies reduce patients' willingness to share information and reduce their positive attitudes toward the technologies [14,17]. In the same vein, reducing privacy concerns makes patients more comfortable in sharing their health information [5,17]. Sometimes, individuals are willing to give up some privacy if they consider mHealth technology as beneficial; however, this depends on the sensitivity of the information they are providing [18]. Due to these privacy concerns, there are many innovations for increasing patient trust and privacy protection [19].

Many researchers have used machine learning methods, including neural networks, random forests, and support vector machines, to predict outcomes such as depression, mood, and stress using mobile data [20-26]. Various physiological and behavioral features help predict these outcomes, such as skin temperature, EDA, 3-axis accelerometer data, mobility, sleep, and self-reported histories [20-23,27,28]. Detecting aberrations in some of these measures can help identify early signs of mental illness [16].

Personalization techniques in machine learning allow parts of a model to be specific to each user while other parts are shared between users, allowing both the sharing of data to estimate certain parameters and fitting to unmeasured covariates of each individual. To an extent, both unique traits among different individuals and commonalities across users are helpful in prediction. Some studies have taken advantage of this and grouped users based on common traits, finding that this improved mood prediction accuracy [8,24,29]. In the same vein, accounting for individual user differences is valuable in prediction using machine learning [8]. There is growing recognition within the mHealth literature [30] that there are trade-offs between individual and centrally trained models and that often there are reasons to make them work in tandem.

In the Methods section, we describe federated learning in the context of mHealth by addressing some of the issues raised in the mHealth literature. Specifically, we discuss the application of federated learning using neural networks to make predictions from mHealth data. We address practical considerations for researchers who consider using federated learning for their study. We then evaluate the performance of federated learning on an mHealth data set to predict subjects' affective states.

Methods

Federated Learning in mHealth

Federated learning is a nascent field within machine learning [31], which arose, at least partially, in response to the opportunities and difficulties presented by personal devices that can record information and perform computation [11]. As such, many practical challenges in mHealth, such as privacy [32,33], user heterogeneity [34], user hardware constraints [35,36], and even incentivizing voluntary participation [37], are already being studied by the machine learning community. The medical science community has also proposed ways to share information while preserving privacy, such as developing models in a distributed manner [38,39], using a federated patient hashing framework for similar patient matching [40], and using collaborative privacy-preserving training to detect protected health information in text [41]. A recent paper describes how federated learning can be useful to various stakeholders in the mHealth field [42].

Federated learning provides a middle ground between extremes in privacy and data utilization. One extreme is collecting data centrally on the server and performing all analyses and training there. The other extreme is individual training, in which each user device trains a completely separate model on its own data.

In centralized learning, we take all individual users' raw data and store them in one location. This unlimited access gives researchers the most flexibility when analyzing the data—clearly any analysis that can be performed with limited data access can be performed with unlimited data access—and the best chance of extracting useful insights from the data. However, it is likely that some individuals are reluctant to have all of their raw data stored in a central location, as the central location could be hacked or because individuals feel that their data are too sensitive to be released. A model trained using federated learning is useful because it is trained without putting all individuals' raw data together in one place. Not only does this

help alleviate concerns about data privacy, but it can also help potential subjects be more willing to engage in studies, as their raw data will not be sent to a foreign central location. As we show in our case study, a model trained in a federated way performs almost as well as a centrally trained model. We refer to the former as the federated model and the latter as the server model.

In individual training, each user device trains a separate model using only their own data; therefore, no information, besides potentially their final model, leaves the user device. We refer to this as the individual model. This provides a very high level of privacy. However, it prevents information sharing across users during model training. Federated learning takes advantage of the information shared across users, which can be especially useful when analyzing health care data. Each person's characteristics and health care need personalization; however, there is also a lot of useful information that can be shared across people, as every person's distribution is not entirely different. In training, the algorithm can learn from similar users and apply this knowledge when predicting another user's outcome. It is significantly easier to use a federated model to make predictions for a new patient because there is a single model that is trained using data from many people, and we do not need previous data from the new patient. In this way, the federated model can do better than the individual model. Thus, federated learning can help improve the prediction accuracy while preserving privacy.

Practical Considerations for Using Federated Learning

The decision to use federated learning must be made when designing the study so that one leaves the data on the user devices and trains the model using federated learning. If data have already been collected from each user device and stored together on a server, privacy has been violated, and we cannot use federated learning to retroactively make the experiment privacy preserving. A small burden might need to be placed on user devices to use federated learning. To minimize disruptive usage of user device bandwidth and compute resources, the local update on the user device should take place while the user device is connected to a nonmetered internet such as Wi-Fi, and the device should be idle and plugged into a charger. To avoid inducing bias in the training process by the overuse or underuse of certain user devices because of their passive adherence to these restrictions, we may need to request that user devices be available for local training during certain times of the day. The structuring of these times should be considered during the experimental design phase. In addition, the server should follow certain protocols to minimize the privacy loss. The model updates may contain some information about the data on the user device; however, the privacy loss from these updates can be mitigated by not storing or viewing any locally updated parameters during each server training round. Similarly, information about when each user device participated in the training should also not be stored.

Although concerns about user privacy deter users from sharing health information, the perceived effectiveness of information security reduces these concerns [5,17]. However, federated learning is not the only tool for privacy, and it does not alone provide perfect privacy. Depending on the privacy requirements of the study, federated learning may or may not be the best method to use. We provide a few examples to illustrate cases where federated learning may not be the correct method.

- Low privacy: If a doctor is collecting data from his or her patients and the patients completely trust their doctor and the security of their server, then there is no need to collect data using federated learning—the doctor can simply take all of his or her patients' raw data. However, the more assured patients are about their information security, the more willing they are to disclose information and the higher their perceived quality of care [17].
- Verifiable privacy: If patients completely distrust the researcher, they may insist on formal guarantees of privacy. This could manifest itself as requiring privacy protection from an honest-but-curious server, wherein the server will attempt to learn anything it can from the information it receives. Under such a requirement, we cannot simply transmit model parameters back to the server, as an individual's model parameters may reveal information about the user. It requires significant technical expertise and engineering effort to implement federated learning under such a strong privacy requirement [43], which may be beyond the resources available in a clinical trial.
- Privacy under model access: Even if no user device data are transmitted to the server, it may be possible to infer information about the data used to train the model, given sufficient access to the model itself, including the reconstruction of specific data points used to train the model [44]. Protection against such attacks is particularly important if the model will be made available to those beyond trusted researchers; in the most extreme cases, researchers themselves may not be trusted. Federated learning provides no guarantees of protection against such attacks; complementary forms of privacy protection, such as differential privacy [45], may be required.

Federated learning is not a silver bullet against privacy loss, and it is not a black-box tool that can be simply tacked onto an existing study. However, when properly integrated into the study design, federated learning can significantly reduce the loss in users' privacy while incurring only a small reduction in model accuracy.

Evaluation of Federated Learning on mHealth Data

To assess the practicality of using federated learning on mHealth data, we compared its predictive performance with that of other predictive machine learning models. We used the WESAD data set for this purpose [9]. We had a 3-class classification problem using physiological and motion data to predict whether each subject was, at the time, performing a task designed to elicit a neutral, stressed, or amused affective state.

We used the WESAD data set because it is publicly available and thus easily accessible by other interested researchers, and it is in the University of California, Irvine, Machine Learning Repository as a data set that measures stress in users using wearables. The data were collected from 15 subjects and contained physiological and motion data measured simultaneously by a wrist device and a chest device during specific tasks designed to capture 3 different affective states:

neutral, stress, and amusement. There was approximately 36 minutes of data for each subject. Using 30-second windows for all 15 subjects, 1087 windows were generated in total. Of these windows, 53.45% (581/1087) were collected during the baseline task, 29.99% (326/1087) during the task designed to elicit stress, and 16.56% (180/1087) during the task designed to elicit amusement.

The authors of the data set found that using physiological and motion data from the chest-worn device was more informative than using physiological and motion data from the wrist-worn device in the 3-class classification problem [9]. The chest-worn and wrist-worn devices differed slightly in the modalities they measured. Thus, we restricted our analysis to the data collected using the chest-worn device. These include electrocardiogram, EDA, electromyogram, respiration, body temperature, and 3-axis acceleration (x-axis, y-axis, and z-axis) measurements collected at 700 Hz. For additional information, we refer interested readers to the original WESAD paper [9].

The authors of the WESAD paper used feature extraction to identify useful features for their predictive models, which often included the mean, SD, minimum, and maximum of measurements. For simplicity, we used these four summary statistics, calculated over 30-second windows, of each of the 8 measurements for each subject as the features in our models. The code to extract our features was derived from Matthew Johnson's GitHub repository [46]. The limitations of our analysis include whether the subject was in the intended affective state and the set of features used.

As we had access to each subject's data, we were able to use a server model. However, in many situations, researchers may not want to access each subject's raw data for privacy reasons. As true federated learning would not store raw data on a central server, we could not use a true federated model and instead use simulated federated learning. Nonetheless, we saw this as a feature and not a limitation. As we had the server data, we could compare the accuracy between our simulated federated model and the truth in the centralized data. If we had implemented true federated learning, we would not be able to compare performance with the centralized, or server, model, or compare performance with the individual model. Much of published federated learning research uses federated learning simulations for experimental results [31,32,34,36].

Data Preprocessing

Of the 15 participants, 12 were male and the remaining 3 were female. Table 1 shows the demographics of the subjects in the WESAD study. Table 2 shows the summary statistics of some of the features used. The full table of summary statistics for the features used is provided in Multimedia Appendix 1.

Table 1. Demographic characteristics of the participants in the study (N=15).

Characteristic	Value, mean (SD)
Age (years)	27.5 (2.4)
Height (cm)	177.6 (6.7)
Weight (kg)	73.1 (10.3)

Table 2. Summary statistics of a subset of features.

Feature	Participants, first quartile	Participants, median	Participants, mean (SD)	Participants, third quartile
Mean EDA[a]	2.0	3.7	4.6 (3.4)	6.3
Mean ECG[b]	7.2E−04	1.1E−03	1.1E−03 (7.8E−04)	1.5E−03
Mean EMG[c]	−3.5E−03	−3.0E−03	−3.0E−03 (9.2E−04)	−2.5E−03
Mean respiration	−0.02	0.05	5.4E−02 (2.0E−01)	0.13
Mean temperature	34	34	34 (1.3)	35
Mean ACC_X[d]	0.73	0.86	8.0E−01 (1.3E−01)	0.90
Mean ACC_Y[e]	−0.06	−0.02	−3.1E−02 (1.0E−01)	0.02
Mean ACC_Z[f]	−0.54	−0.31	−3.5E−01 (2.6E−01)	−0.17

[a]EDA: electrodermal activity.

[b]ECG: electrocardiogram.

[c]EMG: electromyogram.

[d]ACC_X: 3-axis acceleration (x-axis).

[e]ACC_Y: 3-axis acceleration (y-axis).

[f]ACC_Z: 3-axis acceleration (z-axis).

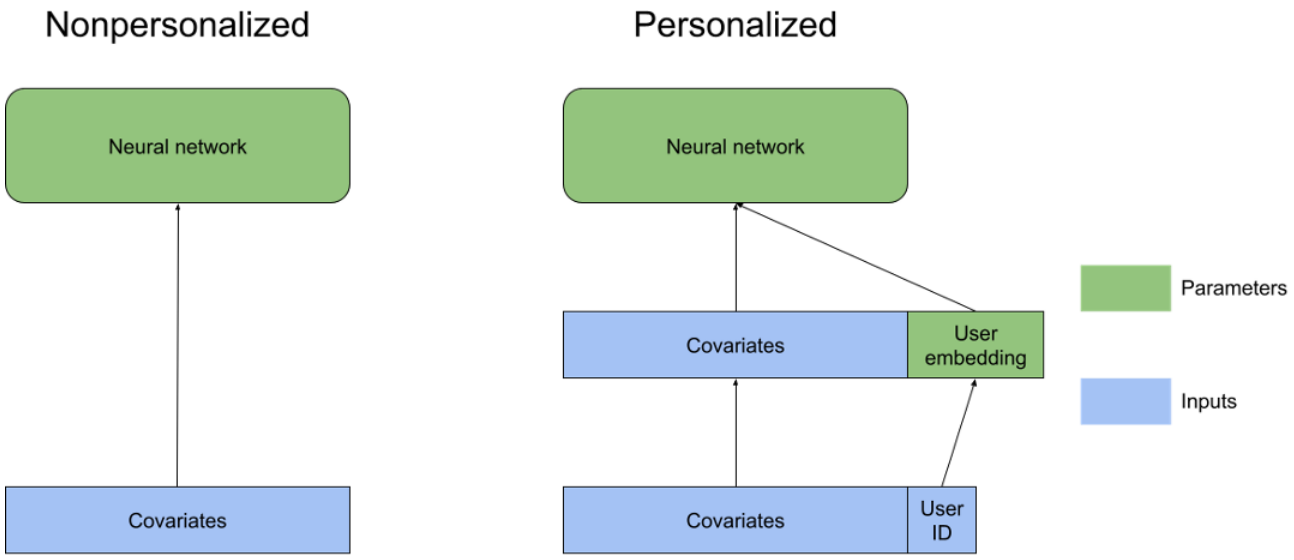
Neural Network Architecture

We tuned the hyperparameters for the individual, federated, and server models separately. We used 3-fold cross-validation to jointly tune the number of epochs and learning rate for the individual and server models. For the federated model, because it involves more hyperparameters, we jointly tuned the number of epochs, learning rate, number of clients sampled per round, number of local updates per round, and step size. The training set consisted of the first two-thirds of the data from each task for each user. The test set consisted of the remaining third of the data from each task for each user. Within each training set, one-third of the training data were used for validation, that is, to tune the hyperparameters.

Our neural networks had 1 dense input layer with 12 hidden units, followed by a dense layer with 10 hidden units, and then followed by a dense layer with 8 hidden units. These layers used the leaky rectified linear unit (leaky ReLU) activation function with a slope of 0.01 where the x-axis is negative. The output layer was a dense layer with 3 hidden nodes and used a softmax activation with categorical cross-entropy loss. We used Adam as the optimizer. We standardized our training data and applied that same standardization to our test set. By privately calculating the mean and SD, standardization can preserve privacy [47]. Our code, which uses Python TensorFlow, and the best hyperparameters chosen by our cross-validation process are available on the authors' GitHub repository [48]. The transformed data are available on GitHub and in Multimedia Appendix 2 [9].

We tested augmentations in the architecture to allow for model personalization. The simplest example of adding personalization to a model is a fixed effects model for linear regression, where each user is given a user-specific intercept, and the other parameters are estimated jointly. This is mathematically equivalent to augmenting each data point by adding a single covariate, where the covariate's value is the same for all data points from the user, and the value is learned by the model-fitting process. We add personalization to our neural network in a similar way, in which we augment each data point by adding a user embedding, with *u* extra covariates that are the same for all data from that user and learned by the model-fitting process (Figure 3). Note that if we solely use user embeddings, then each prediction for the same user would be exactly the same, and the best possible prediction would be the user mean for all data points. This is a very simple example of collaborative personalization [49,50]; in the same way fixed effects models can be extended, there are many extensions of collaborative personalization. However, this simple method is sufficient for our study.

Figure 3. Comparison between the architecture of nonpersonalized and personalized versions of our models. The green blocks represent trained model parameters, and the blue blocks represent the input covariates. Note the user ID is removed from the covariates once it is used to attach the correct parameters, which represent that user's embedding.



User embeddings often encode unmeasured user-specific covariates, meaning that access to user embeddings from a trained model has the potential to expose information about that user. However, a small change in the federated learning protocol can prevent such privacy loss. As user embeddings are user-specific and only updated by training with the user's data, there is no need to transmit gradients for updating any user embeddings to the server. Thus, each person's user embedding may be kept only on that user's device, preventing the server from using that embedding to learn private information about that user.

To choose the dimension of the user embedding for the personalized federated and personalized server models, we used 3-fold cross-validation, as described earlier, with 1D, 2D, and 3D user embeddings. We used the same number of dimensions for both personalized models each time. Using a 2D user embedding achieved the highest accuracy on the validation set for the personalized server model. As such, we used 2D user embeddings in our final personalized federated and personalized server models.

Results

In this section, we present the results for evaluating federated learning on the WESAD data set and compare them with the performance of a server model and an individual model. We include results of personalized federated and server models as well as nonpersonalized federated and server models.

There is inherent variation each time we run a neural network because the initial weights are randomized, and there is randomness in stochastic gradient descent as the neural network tries to find the local minimum. Further, there is variation each time we run a federated model because federated learning takes a random sample of user devices per round of training. As such, we ran each of our models 15 times, using different seed settings each time. We report the median and mean accuracy over 15 tests in Table 3. The distribution of the results over 15 tests and graphs of the average prediction accuracy after each epoch of training of the neural network are presented in Multimedia Appendix 1.

Table 3. Median and mean accuracy of each model over 15 tests.

Model	Accuracy, median	Accuracy, mean (SD)
Personalized server	0.929	0.932 (0.019)
Server	0.897	0.888 (0.028)
Personalized federated	0.929	0.928 (0.018)
Federated	0.853	0.859 (0.021)
Individual	0.899	0.902 (0.021)

The personalized server model achieved the highest accuracy. The personalized federated model came in second, performing nearly as well as the personalized server model. Thus, the personalized server model and the personalized federated model beat their nonpersonalized counterparts.

The individual model outperformed both the nonpersonalized server and nonpersonalized federated models. The individual model is trained separately on each user, whereas the server model takes the data of all individuals at once and cannot adjust for each user. For comparison, a model that always predicts the majority class would attain 53% accuracy on the data set.

These results provide evidence that using personalization, which takes into account individual differences, helped improve the prediction accuracy of the models. This is reasonable for the WESAD data set, considering that each person has varying baseline physiological measurements.

Discussion

Principal Findings

This paper discusses the advantages and challenges of using federated learning as a predictive model in mHealth data collection and demonstrates an empirical example in which federated learning could have been effective. Furthermore, it shows empirical evidence that a federated model can make accurate predictions, is compatible with personalization, and has the added privacy advantage of not storing raw data from individual users. The personalized server model performed the best on the chosen data set, followed closely by the personalized federated model. As a federated model offers more privacy for users than a server model, there is evidence to suggest that federated learning may be a valuable option for collecting and analyzing sensitive mHealth data.

Limitations

The decision of whether to use federated learning is based on a combination of factors, including how much privacy is required, how much data are available, and what resources are available to implement the federated model. Each of the models we tested has advantages and disadvantages. The server model uses all raw data stored in one place for easy access and future use. However, this poses the risk of breaching user privacy if the server is compromised and the risk of the data being accessed and used in a way that the users did not intend. An individual model can maintain user privacy by keeping the data on the user's device and requires less engineering to implement than a federated model. However, the individual model did not perform as well as the personalized server model and personalized federated model in our tests. Nonetheless, if a researcher has a lot of data from each user, using an individual model can be useful for prediction, and a federated model may be unnecessary. The federated model provides more privacy than a server model, as well as reasonably accurate predictions, but it requires some design work before collecting data and software engineering to implement. Moreover, although federated learning provides an added degree of privacy, it does not guarantee privacy, as explained in the Methods section. To achieve stronger privacy guarantees in federated learning, a substantial amount of software engineering is needed.

Federated learning is used for predictive modeling; therefore, implementing it limits the research questions that can be answered from the data. Researchers may often wish to pursue research questions aside from prediction; in the future, it would be interesting to extend federated learning to estimate treatment effects. Other future work includes developing ways to handle missing data for time series in the context of federated learning. In addition, Python TensorFlow has recently released capabilities to apply federated learning in their TensorFlow Federated package, and it would be interesting to compare it with our implementation.

Comparison With Previous Work

Previous research in the mHealth field has explored privacy-preserving methods. Some of these studies do not involve neural networks and federated learning [38-40]. One

study applied a methodology similar to federated learning [41], although it selectively updated parameters from individual users, whereas we used multiple local updates between transmission rounds. As federated learning started in the machine learning community, there have been papers in the computer science and engineering fields regarding collaborative privacy-preserving learning tested on mHealth data [33,51]. Our work bridges the mHealth and machine learning communities by discussing practical considerations for mHealth researchers who want to consider implementing federated learning. Furthermore, we evaluate the effectiveness of federated learning on a publicly available mHealth data set. Our code is available on GitHub to encourage reproducibility of our results.

As mentioned earlier, many mHealth researchers have applied machine learning methods to predict mood and stress [20-23,25,26]. Research has shown that accounting for both unique traits among different individuals and commonalities across users is valuable in prediction [8,24,29]. We demonstrate the use of federated learning as a privacy-preserving method that also has these advantages.

Conclusions

This paper discusses federated learning and its importance in the field of mHealth. For example, federated learning provides added protection to preserve patient confidentiality and inhibits the use of data in a way that the participants in a study did not intend. As federated learning does not store raw data from individual users on a central server, there is no possibility of a central server being hacked and raw data leaked. This provides more privacy to users when recording sensitive data.

Protecting user privacy is critical in mHealth. Having more privacy protection measures in place may encourage people to participate in a study and to be more willing to disclose information that is useful for treatment and research. Federated learning offers additional data privacy and can overcome some common challenges in mHealth data by addressing user heterogeneity and taking advantage of commonalities across users. As such, federated learning has considerable potential to help advance mHealth research.

Acknowledgments

SS was supported by the National Institute of Mental Health (grant number: R01MH101459). AT was supported in part by the National Science Foundation CAREER (grant number: IIS-1452099). JG was supported by the National Science Foundation Research Training Groups in the Mathematical Sciences (grant number: DMS-1646108). JL and JG implemented federated learning and performed data analysis with guidance from AT. JL and JG wrote the Introduction and Methods sections, and JL wrote the Results and Discussion sections. SS provided subject matter expertise and guidance on the methods. All authors have reviewed and revised the manuscript and have read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Federated learning algorithm, summary statistics, and additional figures.
[PDF File (Adobe PDF File), 576 KB - mhealth_v9i3e23728_app1.pdf]

Multimedia Appendix 2
Transformed data.
[ZIP File (Zip Archive), 293 KB - mhealth_v9i3e23728_app2.zip]

References

1. Mechanic D. The functions and limitations of trust in the provision of medical care. J Health Polit Policy Law 1998 Aug;23(4):661-686. [doi: 10.1215/03616878-23-4-661] [Medline: 9718518]
2. Müthing J, Jäschke T, Friedrich CM. Client-Focused Security Assessment of mHealth Apps and Recommended Practices to Prevent or Mitigate Transport Security Issues. JMIR Mhealth Uhealth 2017 Oct 18;5(10):e147 [FREE Full text] [doi: 10.2196/mhealth.7791] [Medline: 29046271]
3. Müthing J, Brüngel R, Friedrich CM. Server-Focused Security Assessment of Mobile Health Apps for Popular Mobile Platforms. J Med Internet Res 2019 Jan 23;21(1):e9818 [FREE Full text] [doi: 10.2196/jmir.9818] [Medline: 30672738]
4. Dehling T, Gao F, Schneider S, Sunyaev A. Exploring the Far Side of Mobile Health: Information Security and Privacy of Mobile Health Apps on iOS and Android. JMIR Mhealth Uhealth 2015;3(1):e8 [FREE Full text] [doi: 10.2196/mhealth.3672] [Medline: 25599627]
5. Damschroder LJ, Pritts JL, Neblo MA, Kalarickal RJ, Creswell JW, Hayward RA. Patients, privacy and trust: patients' willingness to allow researchers to access their medical records. Soc Sci Med 2007 Jan;64(1):223-235 [FREE Full text] [doi: 10.1016/j.socscimed.2006.08.045] [Medline: 17045717]

6. James DC, Harville C2, Whitehead N, Stellefson M, Dodani S, Sears C. Willingness of African American Women to Participate in e-Health/m-Health Research. Telemed J E Health 2016 Mar;22(3):191-197. [doi: 10.1089/tmj.2015.0071] [Medline: 26313323]
7. Ohm P. Broken promises of privacy: Responding to the surprising failure of anonymization. UCLA Law Review 2010;57(6):1701-1778.
8. Taylor SA, Jaques N, Nosakhare E, Sano A, Picard R. Personalized Multitask Learning for Predicting Tomorrow's Mood, Stress, and Health. IEEE Transactions on Affective Computing 2020;11(2):200-213. [doi: 10.1109/TAFFC.2017.2784832]
9. Schmidt P, Reiss A, Duerichen R, Marberger C, Van Laerhoven K. Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection. New York: Association for Computing Machinery; 2018 Presented at: 20th ACM International Conference on Multimodal Interaction; October 16-20, 2018; Boulder, CO, USA p. 400-408 URL: https://archive.ics.uci.edu/ml/datasets/WESAD+%28Wearable+Stress+and+Affect+Detection%29 [doi: 10.1145/3242969.3242985]
10. Hard A, Rao K, Mathews R, Ramaswamy S, Beaufays F, Augenstein S, et al. Federated Learning for Mobile Keyboard Prediction. 2018. URL: https://arxiv.org/abs/1811.03604 [accessed 2021-03-11]
11. Li T, Sahu AK, Talwalkar A, Smith V. Federated Learning: Challenges, Methods, and Future Directions. IEEE Signal Processing Magazine 2020 May;37(3):50-60. [doi: 10.1109/msp.2020.2975749]
12. Yang Q, Liu Y, Chen T, Tong Y. Federated Machine Learning: Concept and Applications. ACM Trans. Intell. Syst. Technol 2019 Feb;10(2). [doi: 10.1145/3298981]
13. Kairouz P, McMahan HB. Advances and Open Problems in Federated Learning. FNT in Machine Learning 2021;14(1). [doi: 10.1561/2200000083]
14. Guo X, Zhang X, Sun Y. The privacy–personalization paradox in mHealth services acceptance of different age groups. Electronic Commerce Research and Applications 2016;16:55-65. [doi: 10.1016/j.elerap.2015.11.001]
15. Perentis C, Vescovi M, Leonardi C, Moiso C, Musolesi M, Pianesi F, et al. Anonymous or Not? Understanding the Factors Affecting Personal Mobile Data Disclosure. ACM Trans. Internet Technol 2017 May;17(2). [doi: 10.1145/3017431]
16. Insel TR. Digital phenotyping: a global tool for psychiatry. World Psychiatry 2018 Oct;17(3):276-277 [FREE Full text] [doi: 10.1002/wps.20550] [Medline: 30192103]
17. Kisekka V, Giboney JS. The Effectiveness of Health Care Information Technologies: Evaluation of Trust, Security Beliefs, and Privacy as Determinants of Health Care Outcomes. J Med Internet Res 2018 Apr 11;20(4):e107 [FREE Full text] [doi: 10.2196/jmir.9014] [Medline: 29643052]
18. Atienza AA, Zarcadoolas C, Vaughon W, Hughes P, Patel V, Chou WY, et al. Consumer Attitudes and Perceptions on mHealth Privacy and Security: Findings From a Mixed-Methods Study. J Health Commun 2015;20(6):673-679. [doi: 10.1080/10810730.2015.1018560] [Medline: 25868685]
19. Boyd AD, Hosner C, Hunscher DA, Athey BD, Clauw DJ, Green LA. An 'Honest Broker' mechanism to maintain privacy for patient care and academic medical research. Int J Med Inform 2007;76(5-6):407-411 [FREE Full text] [doi: 10.1016/j.ijmedinf.2006.09.004] [Medline: 17081800]
20. Umematsu T, Sano A, Taylor S, Picard R. Improving Students' Daily Life Stress Forecasting using LSTM Neural Networks. 2019 Presented at: IEEE EMBS International Conference on Biomedical & Health Informatics (BHI); May 19-22, 2019; Chicago, IL, USA p. 1-4. [doi: 10.1109/bhi.2019.8834624]
21. Pratap A, Atkins DC, Renn BN, Tanana MJ, Mooney SD, Anguera JA, et al. The accuracy of passive phone sensors in predicting daily mood. Depress Anxiety 2019 Jan;36(1):72-81 [FREE Full text] [doi: 10.1002/da.22822] [Medline: 30129691]
22. Suhara Y, Xu Y, Pentland A. DeepMood: Forecasting Depressed Mood Based on Self-Reported Histories via Recurrent Neural Networks. 2017 Presented at: 26th International Conference on World Wide Web; April 2017; Perth, Australia p. 715-724. [doi: 10.1145/3038912.3052676]
23. Spathis D, Servia-Rodriguez S, Farrahi K, Mascolo C, Rentfrow J. Passive mobile sensing and psychological traits for large scale mood prediction. 2019 Presented at: 13th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 2019; Trento, Italy p. 272-281. [doi: 10.1145/3329189.3329213]
24. Mikus A, Hoogendoorn M, Rocha A, Gama J, Ruwaard J, Riper H. Predicting short term mood developments among depressed patients using adherence and ecological momentary assessment data. Internet Interv 2017 Oct 07;12:105-110 [FREE Full text] [doi: 10.1016/j.invent.2017.10.001] [Medline: 30135774]
25. Acikmese Y, Alptekin SE. Prediction of stress levels with LSTM and passive mobile sensors. Procedia Computer Science 2019;159:658-667. [doi: 10.1016/j.procs.2019.09.221]
26. Kornfield R, Sarma PK, Shah DV, McTavish F, Landucci G, Pe-Romashko K, et al. Detecting Recovery Problems Just in Time: Application of Automated Linguistic Analysis and Supervised Machine Learning to an Online Substance Abuse Forum. J Med Internet Res 2018 Jun 12;20(6):e10136 [FREE Full text] [doi: 10.2196/10136] [Medline: 29895517]
27. DeMasi O, Feygin S, Dembo A, Aguilera A, Recht B. Well-Being Tracking via Smartphone-Measured Activity and Sleep: Cohort Study. JMIR Mhealth Uhealth 2017 Oct 05;5(10):e137 [FREE Full text] [doi: 10.2196/mhealth.7820] [Medline: 28982643]
28. NeCamp T, Sen S, Frank E, Walton MA, Ionides EL, Fang Y, et al. Assessing Real-Time Moderation for Developing Adaptive Mobile Health Interventions for Medical Interns: Micro-Randomized Trial. J Med Internet Res 2020 Mar 31;22(3):e15033 [FREE Full text] [doi: 10.2196/15033] [Medline: 32229469]

29. Yu H, Klerman EB, Picard RW, Sano A. Personalized Wellbeing Prediction using Behavioral, Physiological and Weather Data. 2019 Presented at: IEEE EMBS International Conference on Biomedical & Health Informatics (BHI); 2019; Chicago, IL, USA p. 1-4. [doi: 10.1109/bhi.2019.8834456]
30. Schueller SM, Aguilera A, Mohr DC. Ecological momentary interventions for depression and anxiety. Depress Anxiety 2017 Jun;34(6):540-545. [doi: 10.1002/da.22649] [Medline: 28494123]
31. McMahan B, Moore E, Ramage D, Hampson S, Agüera y Arcas B. Communication-Efficient Learning of Deep Networks from Decentralized Data. In: Proceedings of Machine Learning Research. 2017 Presented at: 20th International Conference on Artificial Intelligence and Statistics (AISTATS); Apr 20-22, 2017; Fort Lauderdale, FL, USA p. 1273-1282.
32. Bhowmick A, Duchi J, Freudiger J, Kapoor G, Rogers R. Protection Against Reconstruction and Its Applications in Private Federated Learning. 2019. URL: https://arxiv.org/abs/1812.00984 [accessed 2021-03-11]
33. Liu M, Jiang H, Chen J, Badokhon A, Wei X, Huang M. A Collaborative Privacy-Preserving Deep Learning System in Distributed Mobile Environment. 2016 Presented at: International Conference on Computational Science and Computational Intelligence (CSCI); 2016; Las Vegas, NV, USA p. 192-197. [doi: 10.1109/csci.2016.0043]
34. Zhao Y, Li M, Lai L, Suda N, Civin D, Chandra V. Federated Learning with Non-IID Data. 2018. URL: https://arxiv.org/abs/1806.00582 [accessed 2021-03-11]
35. Wang S, Tuor T, Salonidis T, Leung KK, Makaya C, He T, et al. Adaptive Federated Learning in Resource Constrained Edge Computing Systems. IEEE J. Select. Areas Commun 2019 Jun;37(6):1205-1221. [doi: 10.1109/jsac.2019.2904348]
36. Nishio T, Yonetani R. Client Selection for Federated Learning with Heterogeneous Resources in Mobile Edge. 2019 Presented at: IEEE International Conference on Communications (ICC); 2019; Shanghai, China p. 1-7. [doi: 10.1109/icc.2019.8761315]
37. Kang J, Xiong Z, Niyato D, Yu H, Liang Y, Kim DI. Incentive Design for Efficient Federated Learning in Mobile Networks: A Contract Theory Approach. 2019 Presented at: IEEE VTS Asia Pacific Wireless Communications Symposium (APWCS); 2019; Singapore p. 1-5. [doi: 10.1109/vts-apwcs.2019.8851649]
38. Wu Y, Jiang X, Kim J, Ohno-Machado L. Grid Binary LOgistic REgression (GLORE): building shared models without sharing data. J Am Med Inform Assoc 2012;19(5):758-764 [FREE Full text] [doi: 10.1136/amiajnl-2012-000862] [Medline: 22511014]
39. Lu CL, Wang S, Ji Z, Wu Y, Xiong L, Jiang X, et al. WebDISCO: a web service for distributed cox model learning without patient-level data sharing. J Am Med Inform Assoc 2015 Nov;22(6):1212-1219 [FREE Full text] [doi: 10.1093/jamia/ocv083] [Medline: 26159465]
40. Lee J, Sun J, Wang F, Wang S, Jun CH, Jiang X. Privacy-Preserving Patient Similarity Learning in a Federated Environment: Development and Analysis. JMIR Med Inform 2018 Apr 13;6(2):e20 [FREE Full text] [doi: 10.2196/medinform.7744] [Medline: 29653917]
41. Festag S, Spreckelsen C. Privacy-Preserving Deep Learning for the Detection of Protected Health Information in Real-World Data: Comparative Evaluation. JMIR Form Res 2020 May 05;4(5):e14064 [FREE Full text] [doi: 10.2196/14064] [Medline: 32369025]
42. Rieke N, Hancox J, Li W, Milletarì F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. NPJ Digit Med 2020 Sep 14;3:119 [FREE Full text] [doi: 10.1038/s41746-020-00323-1] [Medline: 33015372]
43. Bonawitz K, Ivanov V, Kreuter B, Marcedone A, McMahan HB, Patel S, et al. Practical Secure Aggregation for Privacy-Preserving Machine Learning. New York: Association for Computing Machinery; 2017 Presented at: ACM SIGSAC Conference on Computer and Communications Security; October 2017; Dallas, TX, USA p. 1175-1191. [doi: 10.1145/3133956.3133982]
44. Yang Z, Zhang J, Chang EC, Liang Z. Neural Network Inversion in Adversarial Setting via Background Knowledge Alignment. New York: Association for Computing Machinery; 2019 Presented at: ACM SIGSAC Conference on Computer and Communications Security; November 2019; London, United Kingdom p. 225-240. [doi: 10.1145/3319535.3354261]
45. Dwork C, Roth A. The Algorithmic Foundations of Differential Privacy. FNT in Theoretical Computer Science 2014;9(3-4):211-407. [doi: 10.1561/0400000042]
46. Johnson M. WESAD. GitHub. 2019. URL: https://github.com/WJMatthew/WESAD [accessed 2021-03-11]
47. Kamath G, Singhal V, Ullman J. Private Mean Estimation of Heavy-Tailed Distributions. In: Proceedings of Machine Learning Research. 2020 Presented at: 33rd Conference on Learning Theory; July 9-12, 2020; Graz, Austria p. 2204-2235. [doi: 10.1007/springerreference_205369]
48. Liu J, Goetz J. Federated Learning on Mobile Health Data. GitHub. 2020. URL: https://github.com/liujessica503/federatedLearning [accessed 2021-03-11]
49. Ni Y, Ou D, Liu S, Li X, Ou W, Zeng A, et al. Perceive Your Users in Depth: Learning Universal User Representations from Multiple E-commerce Tasks. New York: Association for Computing Machinery; 2018 Presented at: 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 2018; London, United Kingdom p. 596-605. [doi: 10.1145/3219819.3219828]
50. Grbovic M, Cheng H. Real-time Personalization using Embeddings for Search Ranking at Airbnb. New York: Association for Computing Machinery; 2018 Presented at: 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; August 2018; London, United Kingdom p. 311-320. [doi: 10.1145/3219819.3219885]

51. Chen Y, Qin X, Wang J, Yu C, Gao W. FedHealth: A Federated Transfer Learning Framework for Wearable Healthcare. IEEE Intell. Syst 2020;35(4):83-93. [doi: 10.1109/mis.2020.2988604]

Abbreviations

EDA: electrodermal activity
mHealth: mobile health
WESAD: Wearable Stress and Affect Detection

Edited by L Buis; submitted 21.08.20; peer-reviewed by L Xiong, C Friedrich; comments to author 18.11.20; revised version received 10.12.20; accepted 25.02.21; published 30.03.21.

Please cite as:
Liu JC, Goetz J, Sen S, Tewari A
Learning From Others Without Sacrificing Privacy: Simulation Comparing Centralized and Federated Machine Learning on Mobile Health Data
JMIR Mhealth Uhealth 2021;9(3):e23728
URL: https://mhealth.jmir.org/2021/3/e23728
doi:10.2196/23728
PMID:33783362

Original Paper

Mobile Apps for Foot Measurement in Pedorthic Practice: Scoping Review

Muhammad Ashad Kabir[1], BSc, MSc, PhD; Sheikh Sowmen Rahman[2], BSc; Mohammad Mainul Islam[1,3], BSc, MSc; Sayed Ahmed[4,5,6,7], BSc, MBA, Cped CM Au; Craig Laird[6,7,8], CPed CM Au

[1]School of Computing and Mathematics, Charles Sturt University, Bathurst, NSW, Australia

[2]Department of Computer Science and Engineering, Chittagong University of Engineering and Technology, Chittagong, Bangladesh

[3]Verizon Media, Sunnyvale, CA, United States

[4]School of Health and Human Sciences, Southern Cross University, Queensland, Australia

[5]Foot Balance Technology Pty Ltd, Westmead, NSW, Australia

[6]Pedi Wiz Digital Technology, Westmead, NSW, Australia

[7]EYM (Ease Your Motion) Pty Ltd, Westmead, NSW, Australia

[8]Walk Easy Pedorthics, Tamworth, NSW, Australia

Corresponding Author:
Muhammad Ashad Kabir, BSc, MSc, PhD
School of Computing and Mathematics
Charles Sturt University
Panorama Ave
Bathurst, NSW, 2795
Australia
Phone: 61 263386259
Email: akabir@csu.edu.au

Abstract

Background: As the use of smartphones increases globally across various fields of research and technology, significant contributions to the sectors related to health, specifically foot health, can be observed. Numerous smartphone apps are now being used for providing accurate information about various foot-related properties. Corresponding to this abundance of foot scanning and measuring apps available in app stores, there is a need for evaluating these apps, as limited information regarding their evidence-based quality is available.

Objective: The aim of this review was to assess the measurement techniques and essential software quality characteristics of mobile foot measurement apps, and to determine their potential as commercial tools used by foot care health professionals, to assist in measuring feet for custom shoes, and for individuals to enhance their awareness of foot health and hygiene to ultimately prevent foot-related problems.

Methods: An electronic search across Android and iOS app stores was performed between July and August 2020 to identify apps related to foot measurement and general foot health. The selected apps were rated by three independent raters, and all discrepancies were resolved by discussion among raters and other investigators. Based on previous work on app rating tools, a modified rating scale tool was devised to rate the selected apps. The internal consistency of the rating tool was tested with a group of three people who rated the selected apps over 2-3 weeks. This scale was then used to produce evaluation scores for the selected foot measurement apps and to assess the interrater reliability.

Results: Evaluation inferences showed that all apps failed to meet even half of the measurement-specific criteria required for the proper manufacturing of custom-made footwear. Only 23% (6/26) of the apps reportedly used external scanners or advanced algorithms to reconstruct 3D models of a user's foot that could possibly be used for ordering custom-made footwear (shoes, insoles/orthoses), and medical casts to fit irregular foot sizes and shapes. The apps had varying levels of performance and usability, although the overall measurement functionality was subpar with a mean of 1.93 out of 5. Apps linked to online shops and stores (shoe recommendation) were assessed to be more usable than other apps but lacked some features (eg, custom shoe sizes and shapes). Overall, the current apps available for foot measurement do not follow any specific guidelines for measurement purposes.

Conclusions: Most commercial apps currently available in app stores are not viable for use as tools in assisting foot care health professionals or individuals to measure their feet for custom-made footwear. Current apps lack software quality characteristics

and need significant improvements to facilitate proper measurement, enhance awareness of foot health, and induce motivation to prevent and cure foot-related problems. Guidelines similar to the essential criteria items introduced in this study need to be developed for future apps aimed at foot measurement for custom-made or individually fitted footwear and to create awareness of foot health.

(JMIR Mhealth Uhealth 2021;9(3):e24202) doi:10.2196/24202

KEYWORDS

foot measurement; foot scanning; mobile app; custom shoes making; apps review; diabetic foot; pedorthics; footcare

Introduction

Background

Poor foot health is often linked to poor performance in both personal and work life [1]. Various anatomical and biochemical factors are responsible for the deterioration of foot health, along with overuse, injury, and external trauma. Maintenance of foot health is necessary to keep humans mobile and independent, and consequential negligence can often cause psychological strain along with physical pain [2].

In an effort to combat problems related to foot health, clinical treatment programs have been widely adopted [3,4]. These programs consist of clinical interventions, and most of the time require clinicians and patients to have regular face-to-face contact for over 1 year. Such interventions have shown variable efficacies due to fluctuations in adherence by patients over time [5-7]. These rigorous health programs can sometimes be time-, resource-, and cost-intensive, and can also be inconvenient for patients given that foot problems increase the possibility of impeding movement capabilities [8,9]. Accordingly, there is a need for novel, low-cost, and widely accessible tools to accurately scan and measure patients' feet, and provide health feedback to patients. This has become a necessity as many patients face significant barriers related to achieving clinical treatments.

The advancement and accessibility of mobile app technology in recent years have enabled efforts to translate the same traditional clinical treatments and intervention programs in the development and growth of the use of foot health mobile apps. The outcome is the development of mobile apps that can provide insight into patients' feet by leveraging the processing capabilities of mobile sensors such as depth-sensing cameras, multicameras, infrared sensors, and features like augmented reality. Many such apps use algorithms and data-mining techniques to suggest foot- and shoe-related solutions based on the procured foot measurement values, whereas others stop at providing basic information such as the suggested size of a shoe based on foot form and suggested forefoot or toe exercises.

Despite an overall increase in app use for foot health conditions, the analysis of several apps belonging to this category has led to the discovery of various problems related to usability, design, and functionality; the limited availability of free apps; the lack of provision for user consent; and, particularly, the lack of certification and the poor quality of the information displayed to achieve their most important goal (ie, to improve patient outcomes) [10-12]. Consequently, these shortcomings have raised questions about the efficacy and applicability of mobile apps used for foot measurement and scanning [10-12].

Objective

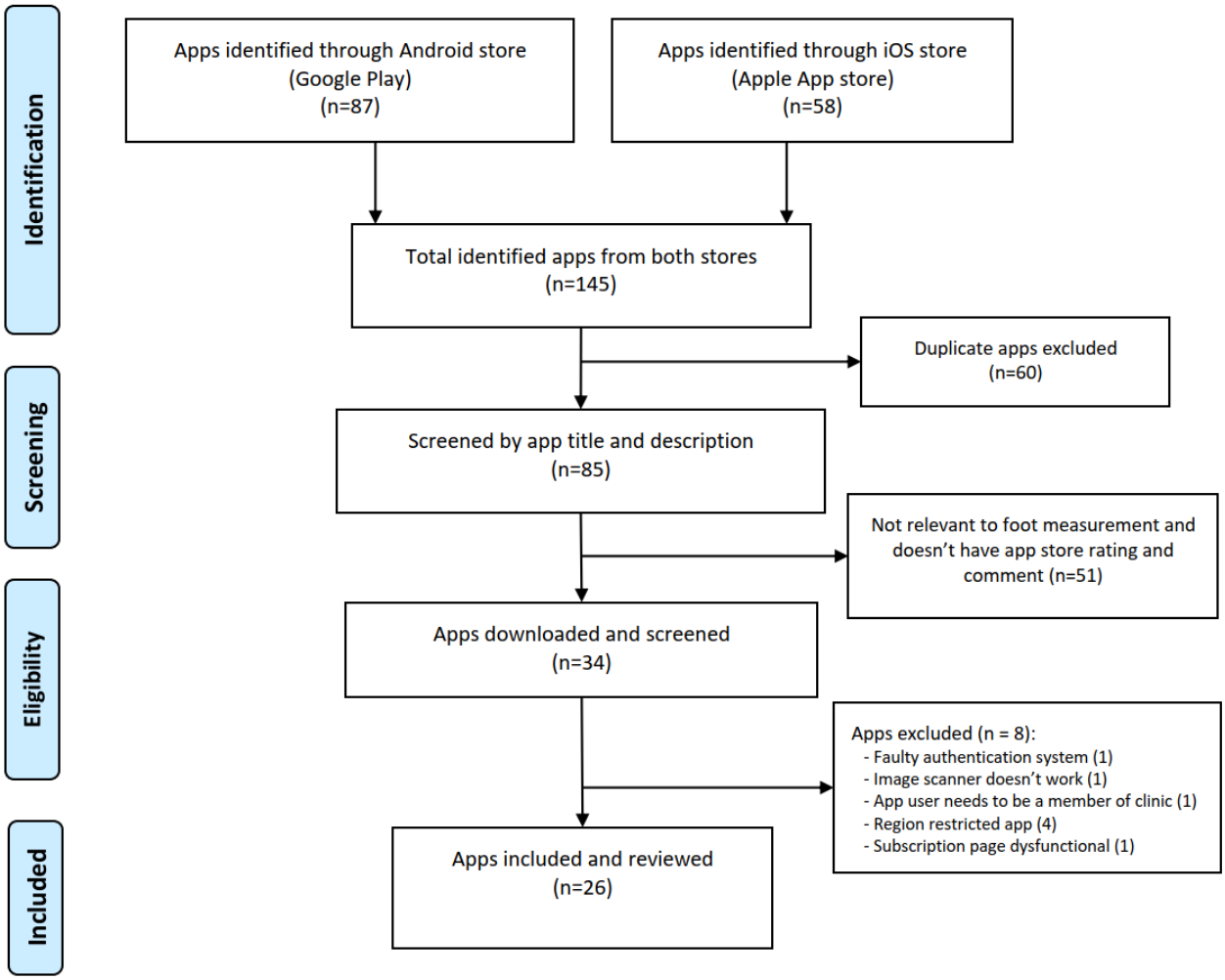
To the best of our knowledge, no study has extensively explored the current scenario of the commercial mobile app market to review and scientifically evaluate apps related to foot measurement. The abundance and rapid growth of such foot measuring apps in app stores, along with the increased adoption rate of these tools by the public, necessitates an assessment of this rapidly growing market. Therefore, the objective of this scoping review was to evaluate the published foot measurement apps in the two major commercial app stores (Apple App Store and Google Play Store). We evaluated the specific criteria of foot properties and the criteria of software quality characteristics, along with the viability of these apps for use as professional tools for foot measurement by pedorthists, podiatrists, orthotists, and individual users. We also investigated the potential of these apps to increase awareness of foot health and foot-related problems.

Methods

App Search Procedure

This review included apps found in the official mobile app stores: Apple App Store and Google Play Store. An electronic search was performed between July and August 2020 in the two app stores for iOS and Android mobile devices, respectively. The search process did not consider any subcategories to which the apps belonged in the app stores. The following terms were used to search for foot measurement apps across the app stores: "foot scanner," "foot app medical," "foot measure," "feet," "measure foot," and "foot length." Region-restricted apps were not considered. The methodology used in this study for the identification, screening, and selection of the apps' matching criteria is displayed in Figure 1.

Figure 1. Flow diagram of study methods.



A secondary search was performed to identify apps that were intended to be used with advanced imaging sensors or other hardware requirements using the following keywords: "foot size," "shoe size," "3D foot," and "foot scan solutions." The results of this latter search were not limited by language, app store description, or rating. No restrictions were imposed in both searches across the different app stores. The search results obtained using the enlisted terms varied significantly between the app stores. As a result, the exact search using the same terms was performed multiple times across different devices to minimize the variance of app indexing and to construct the final inclusion list of foot measurement apps.

The investigation, screening, and extraction phases of the app search and selection processes were performed by the investigators collaboratively. All investigators involved in this operation contributed equally by maintaining a separate list of apps they found within the app stores using the inclusion and exclusion criteria that were finalized for this procedure (see Measures Used in Apps Selection section below). The investigators acted independently on their own smartphone(s) to determine apps fit for selection. During the merging of the app lists, there were conflicts regarding some apps. One such case was an app encountered by one investigator who rejected it from their list, but the same app was included by another investigator. Such cases of discordance were resolved by mutual discussion among the investigators until consensus was reached. The resulting lists of apps were merged, producing the final list of included apps for analysis.

Raters

Three expert raters were selected in line with the proven approach described by Stoyanov et al [13], including (1) a software developer with a Master's degree in software engineering and more than 10 years of mobile app development experience in a renowned company, along with extensive experience using various camera/depth sensors in mobile apps; (2) a computer science graduate with 2 years of mobile app (in particular iOS apps) development experience; and (3) a final-year Bachelor of Computer Science student with 2 years of mobile app development experience. In addition, domain experts (two pedorthists) in the research team trained and educated the raters about the foot measurement processes and the footwear industry.

The raters independently rated all of the apps from the final list of apps compiled by the investigators. Their responses were recorded in a standardized response form (Google Forms) and respective rater response data were extracted from the spreadsheet attached to the form.

Measures Used in Apps Selection

The criteria for the selection of foot measurement apps were based on whether the app serves a purpose that involves measurement of the user's foot. An app was selected for inclusion if it met the following criterion: the app was deemed to be a foot measurement app after screening by title, store description, and store rating relevance. An app was excluded from the study if it met one of the following exclusion criteria: (1) it has neither a star rating nor a user comment on the app store listing page; (2) an identical app from the same developer or publisher is available in both app stores (a duplicated app), and therefore it was excluded from one platform; (3) it is inaccessible/unusable due to region restrictions; and (4) it has been reskinned with a new user interface over an existing app, and hence it is deemed to be the same app in terms of functionality.

Modification of Existing Health App Rating Tools

Selection of Domains and Rating Tool Development

We hypothesized that to properly rate an app with respect to its appropriateness and usability, the app should include multiple features to enable scientifically evaluating its value as a commercially viable foot measurement product. A standardized rating tool would be helpful in this case, especially when there are numerous relevant apps in the app stores. Consequently, an extensive review of prior guidance documents and tools for rating mobile apps was performed to identify the fundamental domains and criteria for determining an app's usefulness and rating. After reviewing several app rating guidance documents and tools, we concluded that each app has its own uniqueness for particular categories of application. We took into consideration prior studies on software quality assessment, and emphasized the key software quality characteristics such as usability, reliability, functionality, and efficiency [14-17]. Our aim was to further build and extend on prior rating tools such as the Mobile App Rating Scale (MARS) [13], end-user version of the MARS (uMARS) [18], and MARS for financial apps (FinMARS) [19], and adapted these to the required evaluation suite for foot measurement apps. With this in mind, we selected the different categorical domains from each rating tool, and proposed an extended version of the mobile app rating tool for foot measurement (FootMARS). This rating tool consists of modifications that we hypothesized to be more important for a foot measuring app. FootMARS considers individual items that were found during the review of the included foot measure apps relevant to the same goal. The finalized rating tool model with the updated overarching domains and individual category items are illustrated in Table 1.

Table 1. Extended foot measurement app rating domains and criteria.

Domain	Criteria
App metadata	App platform
	App store rating
	App store description
	App store URL
	Number of downloads
	Origin
	Developer
App classification	App subcategory
	Applicable age groups
	App price
Aesthetics	Layout consistency and readability
	Content resolution
	Visual appeal
	Group targeting according to app content
General app features	Social sharing feature
	Authentication feature
	User onboarding interfaces
	Content customization
	Visual information
	Data export options
	Subscription options
Performance and efficiency	Bootup efficiency
	Accuracy of features and components
	Responsiveness of app
	Frequency of app crash
	Overheating device issues
	Battery life impact
Usability	Ease of use
	Navigational accuracy
	Gestural design
	Interactivity and user feedback
Measurement-specific functionality	Measurement of foot length and width
	Measurement of foot medial arch height
	Measurement of foot instep or joint girth
	Measurement of short or long heel girth
	Measurement of heel width
	Measurement of shoe size
	Measurement of forefoot tilt/rotation
	Additional setup
	Reconstruction of 3D foot model
	Additional out of scope features
Transparency	User consent
	Accuracy of store description
	Credibility/legitimacy of source
	Feasibility of achieving goals
Subjective quality	Overall star rating
	Overall app purchase preference
	Overall app recommendation
	Frequency of use based on relevance

Domain	Criteria
Perceived impact of app on users	Awareness induction behavior
	Knowledge enhancing behavior
	Scope to improve attitude toward foot health
	Scope to reduce negligence toward foot
	Scope to induce foot-related help-seeking behavior

The key domains essential for evaluating foot measurement apps were identified as follows: app classification, aesthetics, general features, performance and efficiency, usability, measurement-specific functionality, transparency, subjective quality, and the app's perceived impacts on users.

The app quality criteria clustered around the domains, excluding the metadata section, were used to build the app rating scale. Depending on the type of question asked to assess the criteria, each item can score a response as a 5-point Likert scale or a binary response (1 or 5). Cases were found in certain categorical items that were not applicable; thus, the "not applicable" rating was introduced to the scale. Other cases displayed complexity in gaining access to certain types of information, and therefore a rating option "unknown" was added. A few subscale items were excluded from the quantitative measurement since they provided qualitative descriptions about apps that could not be weighed quantitatively. These items are the app metadata domain items, applicable age group item, and app subcategory item.

App Metadata

General information about the apps was abstracted as app metadata from the respective app stores under the app classification category. App metadata include information such as app platform, app URL, store rating, store description, number of downloads, developer information, and origin. However, this information has no impact on the rating scale. App metadata were extracted systematically by two investigators from each app store on a Google Sheet and this dataset was cross-verified by a third investigator for data anomalies.

App Classification

Through an extensive review of prior work on foot measurement and related technologies [20-23], the apps were subcategorized depending on their type of functionality into the following groups: (1) simple size-unit converter, (2) 2D foot scanner, (3) 3D foot scanner, (4) shoe recommender, (5) foot tilt calculator, and (6) foot progress tracker.

Subcategory 1 (simple size-unit converter) apps are the simplest type of apps that take user input values for foot shape and dimensions to produce another category of size or shape, which can be related to both shoe and foot properties. Subcategory 2 (2D foot scanner) apps are more advanced in comparison as they use imaging sensors to acquire 2D data about foot images and use algorithms to compute the user's feet dimensions. The 3D scanner apps (subcategory 3) generally require state-of-the-art techniques and external hardware such as 3D imaging sensors or dimensional digitizing devices. This type of app is generally capable of providing an array of user feet dimensions. The shoe recommender (subcategory 4) and foot tilt calculator (subcategory 5) apps are modified versions of 2D and 3D scanners that do not directly output raw measurement information but rather transform this information into more consumer-friendly, useful views, and derived information. Foot progress trackers (subcategory 6) are common apps that were not specifically designed for applications to feet, but this category was nevertheless included in the study because of the ability of these apps to track measurements of different foot areas over time, either manually or using in-built measurement techniques.

Aesthetics

The current market is dominated by hundreds of thousands of apps that are competing in similar categories with similar functions and outcomes, but are only preferred based on more visually appealing features. Visual appeal is just as important to the success of a commercial product as its core functionality and performance. A proper layout and organization of an app's user interface elements can sometimes be the difference between an app's success and its downfall. This trend is also seen in modern foot scanning apps, which are assessed according to their visual appearance and organization of their layout. Therefore, the key element of a good interface design is to make it clear and simple for users [24]. For measuring the aesthetic properties of an app, the raters considered three factors: (1) the quality of user interface elements in apps, specifically the resolution of images and icons; (2) the overall look and feel of the app's color sets and theme; and (3) the complexity of the contents of the individual screens in the apps.

General Features

Although providing options for as many measurement dimensions as possible is important for a foot measurement app, there are general features that the app must also provide, such as the ability to share and export data to other apps and in various data formats. This feature was selected as a rating item because if it is possible to share data, users are less likely to waste time sharing their foot size information on other platforms, which may be a benefit. An authentication feature is also considered to be a good option. For example, when data are stored against a user's credentials, the data are likely to be stored in the cloud, thus removing the user's dependencies on that particular mobile device on which the app is installed. Over recent years, the importance of content customization and the amount of visual information shown in apps have been found to improve user value; therefore, these features were also included for rating. Additionally, if the app provides subscription packages that may affect user experience in any way, these should also be considered. Negative weights were calculated for rating metrics if the apps had any of the following: faulty authentication system, faulty subscription pages, region

restrictions, and technical issues in image scanning, as crucial app features. All of these points were considered during the rating of apps.

Performance and Efficiency

One of the most important factors contributing to the functionality of an app is its performance score, and how efficiently it can run and provide results on the user's device. The argument is notched up by one point in importance due to the wide scale of customizability of performance components in mobile devices across various global mobile brands. Therefore, as the same foot measurement app may perform differently across different mobile devices with regards to CPU performance, total memory usage, total battery life impact, the possibility of device heating, and other device-specific features, this domain was included as a criterion for rating foot measurement apps, which are known to use significant processing power and complicated processing algorithms to output foot dimensions. To assess the memory usage and battery power consumption of an app, raters tracked the battery usage of apps from the app settings (Android) or auxiliary software (iOS) at intervals of 15 minutes during usage. Raters kept note of performance issues for apps, including (1) whether the app suffered frame drops, (2) the state of the app during any performance throttling, and (3) whether there were any noticeable changes in device temperature during usage.

Usability

The usability of an app refers to the quality of the app's system that is used to achieve the goals of the app. Usability involves the fields of social and behavioral science, and is also linked with the science of design [25]. In prior studies involving human-computer interaction and user-centered design, poor usability and the lack of a proper user-oriented design have been identified as two of the major reasons for low user adoption rates of mobile health (mHealth) apps [26]. Usability testing of a foot scanning app is crucial in determining whether the app has sufficient quality to attract the attention of its target user groups. In today's technology era, when engaging with a mobile app, the user's attention is divided between interaction with the mobile app itself and interaction with the environment [27]. Navigation and ease of use are important measures of app usability since the in-app screen sequence leads the user through different views of the app to obtain the desired information [24]. In prior comparison studies involving comparative app usability testing, it was discovered that the usability of an app in field settings compared to that in laboratory settings varies greatly due to differences in user behavior and the user experience [27]. This signifies that usability testing of an app is a necessary stage in the app development cycle.

Taking into account prior work on app rating systems [13,18,19], we hypothesized that an app's usability could be assessed as "good" given that (1) the app can be operated with ease; (2) the app's navigation flow is uninterrupted; (3) the gestural design (if present in the app) and screen links (buttons, arrows, navigation panels, etc) are consistent across all app pages; and (4) the app provides an interactive experience by taking input from the users and giving feedback when necessary. These conditions were checked by the raters when rating the usability criteria of apps.

Measurement-Specific Functionality

In a foot measurement–specific app, the dimensionality of the features provided by the app is very important. In simple terms, if app A can measure more foot properties than app B, then the potential utility of app A can be considered to be higher than that of app B. Strategizing on this, we decided to include different types of apps that are directly or indirectly involved with the process of foot measurement in this analysis. We reviewed many studies on foot dimension measurements. Currently, foot dimension extraction heavily depends on 3D scanners that are used in both commercial and research areas specifically for measuring foot dimensions [20]. Other studies explored various techniques such as digital light project technology, image sensors, and 3D digitizing devices, including second-generation Kinect, to measure foot length, foot width, and metatarsal/ball girth [21-23]. However, footwear and insole design, instep and medial arch height, ball girth, and forefoot tilt are also necessary foot measurement dimensions [28-30]. In other works, laser scanners were used to scan foot length. The scanned data can also be refined and modified using laser scanners, and can be used for remodeling the human foot [31-33].

Taking these important pedorthic guidelines for foot measurement into account, the app measurement-specific functionality category considered for the weighing of foot measurement apps included the following measurement properties: (1) foot length, (2) foot width, (3) arch height, (4) instep girth, (5) joint girth, (6) short heel girth, (7) long heel girth, (8) heel width, (9) shoe size, and (10) forefoot tilt. Additional discoveries about taking inputs from camera sensors/images, the requirement of calibration markers or extra setup, along with the possibility of reconstruction of a 3D model of the foot were scoped into the rating scheme as functionality subscale items.

Transparency

Mobile apps that use social and personal information for their functioning are common targets for various businesses that capitalize on personalized services [34]. Apps frequently sell private information that is critical to an individual's everyday livelihood without their awareness or approval, and the main cause of this is an improper mobile privacy policy. It must be ensured that when a user gives their consent to private data being accessed by apps, the apps strictly follow specific forms of data protection and regulation rules, and explicitly express to the users how and why their data are being collected, even if users may not understand the direct consequences of this action. Betzing et al [34] examined how increased transparency regarding personal data processing practices in mobile permission requests impacts users in making informed decisions, demonstrating that increasing the transparency of data processing practices increases users' comprehension of their consent decisions. This suggests that obtaining user consent and following data protection rules are important items for the transparency of an app. In the case of foot measurement apps, the above-mentioned constraints should be followed along with

verification of the publisher or developer as to whether the source of the app can be trusted and whether the app is successful in meeting its goals as described in the store description, which are also subject to scrutiny. With this information, a user can make an informed decision before downloading the app by determining the app's authenticity.

To assess the transparency criteria of the selected apps, the following points were considered: (1) providing a general notice or an alert to users before accessing the user's personal information, geolocation information, or private media files; (2) clearly specifying the intention of use of permissions by the app; (3) whether the app store page contained information and a description that aligned with the app goals; (4) whether the app conflicted with the user's activities in unexpected ways; and (5) whether the techniques in the app are feasible to provide the desired output as claimed by the developer.

Subjective Quality

An app's subjective quality refers to its users' key views on the app. These can include personal app ratings, good and bad comments about the app, preference to pay for an app based on its features, and preference to recommend and use an app based on its relevance to the user. Often, users can deduce what an app offers by perusing the reviews from previous users. However, this is subjective, since the general distributed value of app comments and ratings saturate toward an approximated value only as the number of reviews on an app becomes large; thus, this approach to measuring the performance of an app predownload is not effective for apps with few or no user ratings/comments in the app store. However, users often make comparatively in-depth comments on apps with key points that are helpful during the app review phase, which is an optional but valid criterion for apps retaining the saturated direction from user reviews and comments.

We adopted a similar approach for the subjective quality assessment. The app raters answered questions about the degree of satisfaction of use, potential frequency of use in the future, overall app rating, and how likely they were to pay for the apps.

Perceived Impact of App on Users

When an app is downloaded by users, its impact becomes a notable indication of the potential usefulness of the app. Technological developments and ongoing breakthroughs in the fields of computer science and machines have helped thousands of users by guiding their health and preventing death, with a large number of apps designed to improve user health [35]. The main objectives of mHealth apps are to induce awareness about a particular health problem, and increase the user's motivation to avoid and prevent future occurrences related to health. Additionally, mHealth apps may provide intervention techniques and advice useful in decreasing the user's negligence toward their health and increasing help-seeking behaviors targeting solutions to health problems. However, Milne-Ives et al [36] found that most health-oriented apps yield little to no evidence of effectiveness in cases of patient health outcomes and health behavioral changes.

To conclusively support foot measurement apps as useful tools for changing the outcomes of foot health and attention-based behavior, the effectiveness of these apps must be evaluated by app users [36]. Toward this end, the following example strategy was used for assessing the utility of user comments for the rating of apps. The user comments were divided into two types based on the user's review rating of the app: the comment was considered "good" if the rating was 4 stars or above, and otherwise was considered "bad." For iOS apps, the number of downloads could not be viewed publicly and was thus excluded from influencing the assessment of apps with these criteria.

Understanding the impact of an app on users is not directly quantifiable. Therefore, in addition to the approach above, the following strategies were used to assess the extent to which an app was able to make an impact on its users based on their reviews and ratings: (1) how streamlined the app's measurement process was, (2) the reaction of users about certain aspects of the app (store reviews), (3) whether the app had any outstanding features that appealed to the public regarding foot health and foot problems, and (4) whether the app contained information for raising public foot health awareness.

Results

Summary of Search Results

The initial searches using the primary and secondary search terms yielded a set of 145 apps across the two app stores, 41.4% (60/145) of which were excluded from any further review to satisfy the exclusion criteria of app duplication across stores. Of the remaining apps, 35.2% (51/145) were excluded due to failing to satisfy the inclusion criteria of relevance to foot measurement and not having any app store rating or comment. Of the remainder, after installing and using the apps, 5.5% (8/145) were excluded due to various app module errors and store restrictions. In total, 26 apps were selected as eligible to be reviewed as foot measuring apps using our proposed modified rating scale.

Overall Assessment of the Apps

The overall assessment of all eligible apps in this study (reported in Table 2) led to the discovery that most apps belong to the 2D and 3D scanning subcategories. The categorical distribution of all the reviewed apps is shown in Table 3.

Table 2. Assessment scores for foot measurement apps.

App name	Aesthetics	General	Performance	Usability	Functionality	Subjective	Transparency	Impact	Mean (SD)
ShapeCrunch	4.00	3.29	4.33	4.00	2.33	3.75	4.67	4.40	3.85 (.74)
INESCOP YourFeet	5.00	3.00	4.67	4.75	1.73	3.25	5.00	2.60	3.75 (1.26)
FISCHER Scan-Fit	5.00	3.00	4.17	4.00	2.45	3.75	4.67	3.00	3.75 (.88)
Foot Measure	2.50	1.50	3.83	3.75	1.73	1.00	1.33	1.00	2.08 (1.16)
SizeMyShoe	4.25	3.00	4.67	4.00	1.36	2.25	3.67	2.40	3.20 (1.14)
ATLAS-scan your feet!	4.25	3.00	3.33	4.25	1.36	4.00	5.00	2.20	3.42 (1.2)
Jenzy: Easy Kid Shoe Sizing	5.00	3.29	4.67	5.00	1.36	4.25	5.00	2.00	3.82 (1.45)
Shoe Size Meter-foot length	4.25	3.00	4.67	3.75	1.36	3.50	4.33	3.60	3.56 (1.03)
Shoe Size Converter	4.50	1.57	5.00	4.50	1.36	3.50	4.67	2.80	3.49 (1.43)
The Foot Fit Calculator(BikeFit)	5.00	2.14	4.50	4.50	1.67	3.00	4.00	3.60	3.55 (1.19)
Foot Length Converter Size in Lite	4.00	1.5	5.00	4.00	1.36	2.50	4.33	2.20	3.11 (1.39)
myFoot-Rescue your feet!	4.25	4.00	4.67	4.75	3.00	4.00	4.75	4.40	4.23 (.58)
Remeasure Men Body	4.75	4.00	5.00	4.50	1.00	4.75	5.00	3.20	4.03 (1.36)
FootFact	3.25	2.00	4.67	4.25	1.67	1.50	2.33	1.60	2.66 (1.25)
Swift Orthotics	4.50	1.50	4.00	4.25	2.09	4.00	3.00	3.20	3.32 (1.08)
Nimco Professional Shoe Sizing	5.00	4.20	4.50	4.50	4.33	4.50	4.67	3.40	4.39 (0.46)
Shoe-buddy	4.25	3.67	4.17	4.00	2.33	4.00	4.75	3.60	3.85 (0.71)
3D Avatar Feet	4.25	2.14	3.67	4.25	2.67	4.25	4.75	3.20	3.65 (0.91)
SUNfeet	3.75	2.71	4.00	3.75	2.67	3.75	5.00	2.80	3.55 (0.80)
ECLO	5.00	3.86	4.33	4.75	2.33	4.00	5.00	2.40	3.96 (1.07)
FotAppenScan	1.50	1.57	4.50	2.50	1.67	2.25	3.00	2.20	2.40 (0.99)
Ortholutions	4.75	1.50	4.33	4.50	1.67	4.25	5.00	4.20	3.78 (1.38)
AARA Orthotics	4.50	3.50	4.00	4.25	1.67	4.00	4.75	4.40	3.88 (0.97)
Aqualeg	1.75	2.00	3.67	1.50	1.67	2.25	2.00	1.20	2.01 (0.75)
Anodyne Scanner	4.50	2.00	4.00	4.50	1.67	4.25	4.75	3.40	3.63 (1.19)
3DsizeMe	4.75	3.00	4.33	5.00	1.67	4.50	4.75	3.40	3.93 (1.15)

Table 3. Categorical distributions of the reviewed foot measurement apps (N=26).

Subcategory	Count, n (%)
3D Foot Scanner	9 (35)
2D Foot Scanner	8 (31)
Shoe Recommender	4 (15)
Simple Size-unit Converter	2 (8)
Foot Progress Tracker	2 (8)
Foot Tilt Calculator	1 (4)

The 2D and 3D scanning subcategories are the most important since they handle output as raw measurement values, which can be used to make custom-made shoes or to recommend an individualized shoe fit, and to provide users with foot dimension

measurements that may further be used to maintain foot health and prevent foot-related problems. Apps in these two categories mostly use calibration markers such as standard-sized papers (eg, A4, A5) and purchasable barcode stickers. They use the camera sensors on mobile devices to capture, process, and measure users' foot dimensions. Of these apps, 31% (8/26) are the 2D scanning type and 35% (9/26) are the 3D scanning type. Seven of the nine 3D scanning apps require an external sensor (eg, Structure sensor, KinectV2) for the app to function properly and all of these apps can reconstruct a 3D foot model, which is valuable since 3D foot model data are useful for making custom shoes.

Most of the apps targeted a general group of users with no age requirements whereas the apps specifically categorized as 2D and 3D foot scanning apps (17/26, 36%) targeted young adults and adolescents. There were two apps, Jenzy: Easy Kid Shoe Sizing and Remeasure Body, which targeted users in younger age groups (2-18 years old). However, the Remeasure Body app was not a 2D or 3D scanning app but rather a foot progress tracking app. The assessment revealed many commercial apps that use advanced image scanning and processing techniques to determine foot size, shoe type, and size, and even reconstructed 3D foot models to obtain shoe and insole sizes (ATLAS-Scan your feet!, ECLO, Fischer Scan-Fit, Nimco Professional Shoe Sizing, Jenzy: Easy Kid Shoe Sizing, ShapeCrunch).

The assessment results showed that the overall navigability, design, and appeal of current apps scored higher than other aspects of the apps, with usability ranging from 3.75 to 4.50 (Table 2). However, the general range of the measurement-specific functionality was low for all of the apps reviewed, except for Nimco Professional Shoe Sizing. This app performed better in terms of its ability to measure foot length and width, foot instep height, and ball girth. The app also includes a 3D model reconstruction feature of the foot using the procured measurement values. An additional feature of the app is that it can suggest shoe shape and size based on the processed 3D foot model.

Except for one app, all others on the list were free to download, with 31% (8/26) of the apps having subscription packages available; these were essential to obtain full access to the apps' functionality suite (limited access is free).

The mean overall app rating was 3.49 out of 5 (Table 4). Significant differences were detected across domains, most notably impact, general features, and functionality received the lowest ratings. In contrast, the most highly rated domains were performance and transparency. Other domains that scored higher mean values were aesthetics and usability (Table 4).

Table 4. Overall app ratings.

Assessment criteria	Assessment score	
	Mean (SD)	95% CI
Aesthetics	4.17 (0.95)	3.79-4.56
General	2.69 (0.90)	2.33-3.05
Performance	4.33 (0.44)	4.16-4.51
Usability	4.14 (0.74)	3.85-4.44
Functionality	1.93 (0.70)	1.65-2.21
Subjective	3.50 (0.97)	3.11-3.89
Transparency	4.24 (1.04)	3.82-4.66
Impact	2.94 (0.94)	2.56-3.32
Total	3.49 (0.61)	3.25-3.74

Internal Consistency of the Modified Scale and Interrater Reliability

We used Cronbach α [37] to calculate the internal consistencies of the overarching domains of the modified rating scale: aesthetics, general app features, performance and efficiency, usability, measurement-specific functionality, transparency, subjective quality, and perceived impacts on users. For this work, one of the foot measurement apps reviewed in this study, INESCOP YourFeet, was additionally used to rate the internal consistency of our modified rating scale. All three independent raters considered the internal consistency of all subscale items as "good" to "high." The overall internal consistency of the modified rating scale was high at α=.84, which is considered excellent (eg, [38]). The Cronbach α values were also in the range of good to excellent for the other subscales: aesthetics (α=.88), general features (α=.92), performance (α=.70), usability (α=.84), functionality (α=.92), subjective (α=.79), transparency (α=.90), and perceived impacts (α=.78).

The same method was used to measure the interrater reliability of our raters who independently reviewed the set of 26 apps in this study. The interrater agreement score ranged between .54 and .70, which is considered as a fair to good level of rater reliability or agreement [38].

Comparison of Store Ratings and Rating Scales

The store ratings of the reviewed apps were compared to the score of the apps from our rating scale (Table 5). The standard deviation of the difference in the two scores for the reviewed apps was 1.07. This deviation is not too poor considering that the score in our rating scale is an aggregated mean of the various domains that are necessary to specify the quality and criteria of

foot measuring apps. Even though the ratings were within a close range of spread, there is a need for more than just the store rating to motivate users to install and use the apps since the apps reviewed in this study are not intended to be used as casual apps (for evaluation of the study's objective). During the review, no star ratings for the apps SUNFeet, Ortholutions, AARA Orthotics, and Aqualeg were found, and thus these apps were excluded from the calculation of the total standard deviation.

Table 5. Comparison of app ratings from the app stores and the developed rating scale.

App name	Measured score	App store rating
ShapeCrunch	3.85	4.50
INESCOP YourFeet	3.75	4.10
FISCHER Scan-Fit	3.76	2.10
Foot Measure	2.08	2.00
SizeMyShoe	3.20	2.10
ATLAS-scan your feet!	3.42	2.00
Jenzy: Easy Kid Shoe Sizing	3.82	3.30
Shoe Size Meter-foot length	3.56	3.70
Shoe Size Converter	3.49	4.90
The Foot Fit Calculator (BikeFit)	3.55	3.00
Foot Length Converter Size in Lite	3.11	3.00
myFoot-Rescue your feet!	4.23	3.40
Remeasure Men Body	4.03	4.40
FootFact	2.66	3.10
Swift Orthotics	3.32	5.00
Nimco Professional Shoe Sizing	4.39	2.00
Shoe-buddy	3.85	3.30
3D Avatar Feet	3.65	4.50
ECLO	3.96	5.00
FotAppenScan	2.40	3.50
Anodyne Scanner	3.63	5.00
3DsizeMe	3.93	4.20

Measurement Criteria for Apps

The key function of foot measurement apps is the measurement of foot dimensions so that the users are aware of the current condition of their feet. In accordance with the selected measurement criteria for foot measurement apps, we evaluated the comprehensiveness of the reviewed apps regarding the criteria. As shown in Table 6, there was substantial variation in the number of foot dimensions measured by the apps. Table 6 provides a transformed view of the number of measurement items of apps as discussed in the Measurement-Specific Functionality subsection of the Methods section.

Table 6. Assessment criteria for the measurement functionality of apps.

Measurement criteria	Google Play Store (n=11), n (%)	Apple App Store (n=15), n (%)	Total (N=26), n (%)
Foot length	6 (55)	7 (47)	13 (50)
Foot width	6 (55)	5 (33)	11 (42)
Foot arch (medial) height	1 (9)	2 (13)	3 (12)
Foot instep girth	0 (0)	1 (7)	1 (4)
Foot joint girth (ball girth)	0 (0)	3 (20)	3 (12)
Short heel girth	0 (0)	0 (0)	0 (0)
Long heel girth	0 (0)	0 (0)	0 (0)
Heel width	0 (0)	2 (13)	2 (8)
Shoe size	2 (18)	2 (13)	4 (15)
Forefoot tilt	0 (0)	2 (13)	2 (8)
3D foot model reconstruction	1 (9)	10 (67)	11 (42)

The general distribution of the measured foot dimensions of the reviewed apps was very low overall. From the total of 26 apps, only 50% can measure foot length, 42% can measure foot width, 15% can determine shoe size for users, 12% can measure foot instep height and ball girth, and 4% can measure foot instep girth and heel width. Important dimensions that are crucial for the construction of shoes, insoles, and foot wedges (eg, for bicycle pedals, saddles) were missing from the measurement features offered by all of the reviewed apps. Most of the reviewed apps from Google Play Store were poor compared to those available from Apple App Store. However, since apps were excluded based on duplication across app stores and region-restriction criteria, this information should not be used to reflect the current state of platform-specific apps (ie, Android devices have a lower variety of apps devised for foot measurement compared to iOS devices).

Further evaluation of the measurement criteria showed that the maximum number of dimensions measured by a foot measuring app was five. No app measured more than half of the required foot measurement criteria, and half of the apps (50%, 13/26) only measured one dimension. There were 3 (12%) apps that measured five dimensions, 3 (12%) apps that measured four dimensions, 1 (4%) app that measured three dimensions, 3 (12%) apps that measured two dimensions, and 3 (12%) apps that measured no dimensions at all. In general, most of the reviewed apps were not suitable for foot measurement in clinical practice for custom-made shoes and insoles/orthoses or for general use by the public.

User Reviews from App Store

User reviews provide a rich source of information about the performance and future viability of apps. Hence, to drive the commercial success of apps (in any category), developers and publishers aim to receive good and positive reviews as these are a crowdsourced quality indicator of apps [39].

The analysis of user comments regarding the foot measurement apps from both stores showed that most of the apps lack good feedback or positive reviews from app users, since the number of downloads of the related apps was generally low and an estimated 31% (8/26) of the apps had only 100 to 1000 users. The number of downloads of iOS apps could not be viewed and there were too few reviews with information from users for proper assessment of these apps. These findings are similar to those of Vasa et al [39], who indicated that users tend to write little to nil when reviewing an app that performs better in its category. We found that 42% (11/26) of the apps had very short or no informative reviews to support their ratings.

However, 54% (14/26) of the apps had some informative reviews in the app stores, which were contrasted with their ratings/scores. One of the most positive sets of user responses was found for the app ShapeCrunch. This app is a foot scanner that uses machine learning–generated 3D printed insoles to reduce the discomfort of foot pain. Based on the reviews of several users, this app is rated excellent on the grounds that it can provide properly fitted insoles for users with a wide variety of foot problems. The app Swift Orthotics also received a good review, which was consistent with the overall ratings both in the app store and in the modified evaluation scale. This app uses augmented reality technology to measure foot dimensions, and was the only other app that could measure heel width along with Fischer Scan-Fit among the reviewed apps. Strong positive feedback with no mention of drawbacks was received for the apps ECLO and 3DSizeMe, which were positively correlated with the ratings of the independent raters, although ECLO is a shoe recommendation app that can order shoes based on user scans from the ECLO online store. The app 3DSizeME has great support for Structure Sensor, an advanced 3D scanner, and there are many companies that can directly support 3DSizeMe file formats to order custom shoes and insoles. By contrast, apps such as Aqualeg, Ortholutions, and AARA Orthotics, which use external scanning mechanisms, did not receive much public exposure although they scored high ratings according to the devised rating scale, and all of these apps had features for visualizing and either uploading foot scans to company servers or exporting the model data to order shoes, insoles, and even medical casts for different body parts. Similar results were found for the 2D scanning app INESCOP YourFeet, which did not have good store descriptions, but the store ratings were consistent with our measured ratings. However, this app lacked good functionality in terms of providing detailed foot

information to make custom footwear (although it can measure regular shoe size).

Our raters found inconsistencies between app user reviews and the actual performance of some apps: Shoe Size Meter-foot length, SizeMyShoe, FISCHER Scan-Fit, Jenzy: Easy Kid Shoe Sizing, INESCOP YourFeet, and The FootFit Calculator. Two of the apps that performed better than the rest of the 3D scanner apps reviewed were SUNFeet and 3D Avatar Feet; however, their respective app store pages had no user reviews at all.

From the above analysis of foot measurement app user reviews, we highlight the inconsistency of rater app ratings in this study and the app store user reviews. This may be because the performance and functionality of apps are dependent on the specific device version and software being used. This suggests that most of the currently available apps suffer from device architecture and build-related problems, and need to be further optimized. Another possible cause of this rating versus review inconsistency is that when users are rating apps on the store, they generally do not focus on domains such as perceived impacts, transparency, and the technical functionality of the apps, and thus their comments are not representative of all of the apps' features.

Discussion

Principal Findings

The findings of this review are discussed from three main aspects: (1) the viability of apps in podiatric practice for making custom shoes, (2) the viability of apps for individual use for general measurement purposes, and (3) the potential of inducing behavioral changes about foot health and foot-related problems among users.

This review demonstrates that although there are a handful of foot measurement apps available in commercial stores, the performance of the apps with regard to the objective(s) of this study is poor and the features are insufficient. The objective of foot measuring apps available in app stores to be used in clinical practice to facilitate custom-made shoes is not being achieved using the current configuration. Although apps may be used for different purposes such as online shopping for shoes and insoles, and the casual measurement of feet, they are deemed unusable as pedorthic tools for the professional measurement of foot dimensions, which require comprehensive fulfillment of the measurement criteria determined by this study to achieve proper and precise measurements for custom-made shoes and insoles.

Most of the reviewed apps met less than half of the measurement dimension criteria required to properly measure the feet of users. Some of the apps that were used for foot measurements did not provide sufficient relevant information about the actual foot dimensions they measured; rather, this information was used and processed for other purposes. No app included any information about the degree of accuracy that could be achieved from the measurement with the used technologies.

In general, most of the current apps belonging to 3D categories could reconstruct a 3D model of the feet. Some apps measure a small number of foot properties and send scanned information to their internal servers to construct a 3D foot model. Other apps take in information to find and fit custom-made shoes and insoles/orthoses for users (possibly with foot-related disabilities and zonal pain), and have them delivered to the user's home. However, except for 6 (23%) apps (Nimco Professional Shoe Sizing, Fischer Scan-Fit, 3D Avatar Feet, SUNFeet, 3DSizeME, and Anodyne Scanner), no other app reported the user's foot dimension information with sufficient detail to enable custom-made shoes (and insoles) to be built. The app 3DSizeME is mentionable in this respect owing to its performance and data shareability options, which are lacking in most current apps, thus making the mobility of data difficult. In general, the apps tended to focus on particular measurement criteria, which cannot completely describe the structure of the foot to make custom-made shoes.

With respect to the general measurement use of individuals, most of the reviewed apps were either developed for commercial purposes or were most likely to be used as lookup apps to obtain preferred foot/shoe sizes. However, the usage frequency of apps underperforms by a large margin; 65% (17/26) of the apps subjectively reviewed by raters had an average subjective usage frequency of under 10 times over 12 months, 31% (8/26) of the apps had a usage frequency of 10 to 50, and only one app was properly usable with a usage frequency ≥50 over 12 months. Apps in the shopping category that met this usage range were ATLAS-Scan your feet!, Shoe Size Converter, and Nimco Professional Shoe Sizing. The usage statistics and the calculated ratings of such apps in this review, together with analysis of the user reviews from the app stores, suggest that the cause of the poor usage of apps, even though a large percentage of these apps provided all of the included features for free, was a lack of more features and the presence of poorly optimized features in the apps. Thus, for individual use, the apps are not likely to be used as foot measurement tools of high value. Some apps had detailed complaints about their inability to properly size feet and shoes, while others suffered from performance issues, including battery draining, network errors, and overheating problems with prolonged use. In some apps, the structural flow of technical aspects was confusing and difficult for the users to follow.

This review showed that the apps that fell into the shopping category performed better in calculating foot measurements compared with the foot-scanning apps that output raw measurement values. Seven out of 26 apps met this criterion, and all apps had consistently good mean scores.

Based on the accuracy of their store descriptions and the credibility of the app developers, the apps were mostly consistent about their intended use; however, some apps were not explicit about the consent to the use of data. Some apps displayed warning dialogs to notify users about providing private data access, whereas some apps did not offer sufficient information about their policy on protecting the personal and private information of users on their support websites or on the app privacy policy page (if relevant), which raises a question about their intention to ensure the privacy of user data.

The results also showed that current apps have low fidelity for inducing behavioral changes to promote foot health and an

awareness of foot-related problems. The apps mostly did not meet the required basic measurement dimensions needed to properly measure feet, and were therefore deemed to be unsuitable for inducing health-related awareness (79.3% of app responses were negative) and help-seeking behavior (82% of app responses were negative) related to foot health and problems. However, this study showed that most of the apps used in cases other than the direct measurement of foot dimensions had a positive effect (64.7% positive responses) on the necessity of buying properly sized shoes and insoles. Such impacts suggest that although the commercial market has grown with the increase in foot measurement–related apps, the technical quality of the apps needs further improvement. Developers should carefully follow guidelines such as those provided by our rating scheme when developing apps that are publicly available for making custom-made footwear or individualized shoe fit for all types of feet.

Strengths and Limitations

The search methods that were used in this study followed a modeling pattern similar to that of previous studies involving various mHealth categories, including the management of various mental and chronic diseases [26], diet [10], drugs and alcohol [40], physical activity [36], weight [41], and diabetes [11,12]. This work focused on apps that were not access-restricted by region. Various domains of software characteristics hypothesized to be important for foot measurement were presented and a rating scale was created based on thorough reviews of published foot measurement apps. Although these apps were not verified for the presence of applied knowledge of foot morphology and foot health, a general impact value of the apps was taken to assess how the consumers viewed the features provided by these apps and whether they could produce a sense of awareness about foot health. It should be noted that despite raters independently testing all of the selected and eligible apps for the review, these values should not be interpreted to focus on any particular criteria or item, as this was not the main objective of the study. The findings presented are a broad characterization of the general quality characteristics and measurement-specific features that should be presented in foot measurement apps to create custom footwear products and for general use by individuals to enhance awareness of foot health, and thus represent the current state of the commercial app market with regard to foot measurement apps. This study provides insight into the technical knowledge of developers and publishers about foot health and foot measurement, and points to possible directions to advance research and development in the foot health and measurement categories.

Future Directions

Although we aimed to assess all of the important and relevant software characteristics of apps, certain assessment criteria such as the accuracy of the measured dimensions of feet, CPU, memory, and battery usage could not be assessed accurately due to resource constraints. Hence, future research will focus on determining foot measurement accuracy and introduce more app stores for an even more comprehensive review. Future work will also investigate the possibility of including more features as part of the categorical criteria to ensure more robust foot measuring apps with enhanced technical features. Another focus of this research would be a direct extension and update of the review since, among the vast collection of foot measurement apps reviewed for this study, we may have missed apps because of our search criteria, and there may be apps that were not available due to the regional restriction of the app stores.

Conclusion

The conclusion drawn from this review on foot measurement apps is that most mobile apps in the app stores do not sufficiently meet the criteria required to manufacture custom-made footwear or recommend an individualized shoe fit. Although a few apps provide some information about the user's feet, they require users to enter the data manually. We believe that only by addressing the entirety of issues found and by applying more caution to implementing features that are required for the completeness of foot measurement will developers be able to bring useful apps to app stores that are suitable for direct professional use in clinical practice for custom-made footwear. Such apps will also motivate individual users to properly address foot problems and become more aware of the importance of foot health.

Acknowledgments

The authors would like to thank Kazi Miftahul Hoque, Sabiha Samad, and Sadman Sakib Chowdhury for their help with the app ratings. The work was partially funded by a Foot Balance Technology Pty Ltd grant (Project number: RM102978).

Conflicts of Interest

None declared.

References

1. Why Preventative Foot Care is Important. Southeast Orthopedic Specialists. 2016. URL: https://se-ortho.com/foot-care-important/ [accessed 2021-02-23]
2. Gray A. A strong base: the importance of foot health. The Pharmaceutical Journal. 2017 Aug 17. URL: https://pharmaceutical-journal.com/article/feature/a-strong-base-the-importance-of-foot-health [accessed 2021-02-23]
3. Riskowski JL, Hagedorn TJ, Hannan MT. Measures of foot function, foot health, and foot pain: American Academy of Orthopedic Surgeons Lower Limb Outcomes Assessment: Foot and Ankle Module (AAOS-FAM), Bristol Foot Score (BFS), Revised Foot Function Index (FFI-R), Foot Health Status Questionnaire (FHSQ), Manchester Foot Pain and Disability

Index (MFPDI), Podiatric Health Questionnaire (PHQ), and Rowan Foot Pain Assessment (ROFPAQ). Arthritis Care Res (Hoboken) 2011 Nov;63(Suppl 11):S229-S239. [doi: 10.1002/acr.20554] [Medline: 22588747]
4. Edwards K, Borthwick A, McCulloch L, Redmond A, Pinedo-Villanueva R, Prieto-Alhambra D, et al. Evidence for current recommendations concerning the management of foot health for people with chronic long-term conditions: a systematic review. J Foot Ankle Res 2017;10:51 [FREE Full text] [doi: 10.1186/s13047-017-0232-3] [Medline: 29201147]
5. Williams AE, Davies S, Graham A, Dagg A, Longrigg K, Lyons C, North West Clinical Effectiveness Group for the Foot in Rheumatic Diseases (NWCEG). Guidelines for the management of the foot health problems associated with rheumatoid arthritis. Musculoskeletal Care 2011 Jun;9(2):86-92. [doi: 10.1002/msc.200] [Medline: 21259413]
6. Janisse D, Janisse E. Pedorthic management of the diabetic foot. Prosthet Orthot Int 2015 Feb;39(1):40-47. [doi: 10.1177/0309364614535233] [Medline: 25614500]
7. Ahmed S, Barwick A, Butterworth P, Nancarrow S. Footwear and insole design features that reduce neuropathic plantar forefoot ulcer risk in people with diabetes: a systematic literature review. J Foot Ankle Res 2020 Jun 04;13(1):30 [FREE Full text] [doi: 10.1186/s13047-020-00400-4] [Medline: 32498719]
8. López-López D, Becerro-de-Bengoa-Vallejo R, Losa-Iglesias ME, Palomo-López P, Rodríguez-Sanz D, Brandariz-Pereira JM, et al. Evaluation of foot health related quality of life in individuals with foot problems by gender: a cross-sectional comparative analysis study. BMJ Open 2018 Oct 18;8(10):e023980 [FREE Full text] [doi: 10.1136/bmjopen-2018-023980] [Medline: 30341140]
9. Boulton AJM, Vileikyte L, Ragnarson-Tennvall G, Apelqvist J. The global burden of diabetic foot disease. Lancet 2005 Nov 12;366(9498):1719-1724. [doi: 10.1016/S0140-6736(05)67698-2] [Medline: 16291066]
10. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
11. Bonoto BC, de Araújo VE, Godói IP, de Lemos LLP, Godman B, Bennie M, et al. Efficacy of mobile apps to support the care of patients with diabetes mellitus: a systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2017 Mar 01;5(3):e4 [FREE Full text] [doi: 10.2196/mhealth.6309] [Medline: 28249834]
12. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016 May 16;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
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 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
14. Faux F, Bastide R, Souf N, Zgheib R. Smartphone-Based Collaborative System for Wounds Tracking. 2016 Presented at: 8th International Conference on e-Health, Telemedicine and Social Medicine (eTELEMED); April 24, 2016; Venice, Italy p. 104-109 URL: https://hal.archives-ouvertes.fr/hal-01523684/
15. Poon TWK, Friesen MR. Algorithms for size and color detection of smartphone images of chronic wounds for healthcare applications. IEEE Access 2015;3:1799-1808 [FREE Full text] [doi: 10.1109/ACCESS.2015.2487859]
16. Friesen M, Hamel C, McLeod R. A mHealth application for chronic wound care: findings of a user trial. Int J Environ Res Public Health 2013 Nov 19;10(11):6199-6214 [FREE Full text] [doi: 10.3390/ijerph10116199] [Medline: 24256739]
17. Vos-Draper T. Poster 29: Wireless seat interface pressure mapping on a smartphone: feasibility study in users with SCI. Arch Phys Med Rehab 2013 Oct;94(10):e21-e22. [doi: 10.1016/j.apmr.2013.08.075]
18. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
19. Huebner J, Schmid C, Bouguerra M, Kunze A. FinMARS: A mobile app rating scale for finance apps. 2019 Presented at: Proceedings of the 9th International Conference on Information Communication and Management (ICICM); 2019; Prague, Czech Republic p. 6-11 URL: https://dl.acm.org/doi/10.1145/3357419.3357428
20. Telfer S, Woodburn J. The use of 3D surface scanning for the measurement and assessment of the human foot. J Foot Ankle Res 2010 Sep 05;3:19 [FREE Full text] [doi: 10.1186/1757-1146-3-19] [Medline: 20815914]
21. Geng J. Structured-light 3D surface imaging: a tutorial. Adv Opt Photon 2011 Mar 31;3(2):128. [doi: 10.1364/aop.3.000128]
22. Liu W. Accuracy and reliability of a technique for quantifying foot shape, dimensions and structural characteristics. Ergonomics 2010 Nov 10;42(2):346-358. [doi: 10.1080/001401399185702]
23. Wang M, Wang X, Fan Z, Zhang S, Peng C, Liu Z. A 3D foot shape feature parameter measurement algorithm based on Kinect2. EURASIP J Image Video Proc 2018 Nov 6;2018(1):119. [doi: 10.1186/s13640-018-0368-5]
24. Georgieva ES, Smrikarov AS, Georgiev TS. Evaluation of mobile learning system. : Procedia Computer Science, Sciencedirect; 2011 Presented at: World Conference on Information Technology (WCIT); October 6-10, 2010; Istanbul, Turkey p. 632-637. [doi: 10.1016/j.procs.2010.12.106]
25. Lindgaard G. Early traces of usability as a science and as a profession. Interact Comput 2009 Dec;21(5-6):350-352. [doi: 10.1016/j.intcom.2009.03.006]

26. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
27. Kaikkonen A, Kekäläinen A, Cankar M, Kallio T, Kankainen A. Usability Testing of Mobile Applications: A Comparison between Laboratory and Field Testing. Journal of Usability Studies 2005;1(1):4-16 [FREE Full text]
28. Williams A, Nester C. Modern footwear. In: Pocket Podiatry: Footwear and Foot Orthoses. London, UK: Elsevier; 2010:69-80.
29. Pivecka J, Laure S. The Shoe Last: Practical Handbook for shoe Designers. Slavicin, Czech Republic: Pivecka Jan Foundation; 1995.
30. Baba K. Foot measurement for shoe construction with reference to the relationship between foot length, foot breadth, and ball girth. J Hum Ergol (Tokyo) 1974 Dec;3(2):149-156. [Medline: 4465404]
31. Carocci M, Lazzari S, Rodella R, Sansoni G. 3D range optical sensor: analysis of the measurement errors and development of procedures for their compensation. : Society of Photo-Optical Instrumentation Engineers (SPIE); 1998 Presented at: Proceedings of Three-Dimensional Image Capture and Applications; March 1998; San Jose, CA URL: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/3313/1/3D-range-optical-sensor--analysis-of-the-measurement-errors/10.1117/12.302452.short [doi: 10.1117/12.302452]
32. Luximon A, Goonetilleke RS. Foot shape modeling. Hum Factors 2004;46(2):304-315. [doi: 10.1518/hfes.46.2.304.37346] [Medline: 15359679]
33. Li P, Corner B, Paquette S. Segmenting 3D surface scan data of the human body by 2D projection. : Society of Photo-Optical Instrumentation Engineers (SPIE); 2000 Presented at: Proceedings of Three-Dimensional Image Capture and Applications; March 2000; San Jose, CA. [doi: 10.1117/12.380038]
34. Betzing JH, Tietz M, vom Brocke J, Becker J. The impact of transparency on mobile privacy decision making. Electron Markets 2019 Feb 07;30(3):607-625. [doi: 10.1007/s12525-019-00332-3]
35. McKay FH, Wright A, Shill J, Stephens H, Uccellini M. Using health and well-being apps for behavior change: a systematic search and rating of apps. JMIR Mhealth Uhealth 2019 Jul 04;7(7):e11926 [FREE Full text] [doi: 10.2196/11926] [Medline: 31274112]
36. Milne-Ives M, Lam C, De Cock C, Van Velthoven MH, Meinert E. Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: systematic review. JMIR Mhealth Uhealth 2020 Mar 18;8(3):e17046 [FREE Full text] [doi: 10.2196/17046] [Medline: 32186518]
37. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951 Sep;16(3):297-334. [doi: 10.1007/BF02310555]
38. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess 1994 Dec;6(4):284-290. [doi: 10.1037/1040-3590.6.4.284]
39. Vasa R, Hoon L, Mouzakis K, Noguchi A. A preliminary analysis of mobile app user reviews. New York, USA: ACM; 2012 Nov Presented at: 24th Australian Computer-Human Interaction Conference (OzCHI); November 2012; Melbourne Australia p. 241-244. [doi: 10.1145/2414536.2414577]
40. Tofighi B, Chemi C, Ruiz-Valcarcel J, Hein P, Hu L. Smartphone apps targeting alcohol and illicit substance use: systematic search in commercial app stores and critical content analysis. JMIR Mhealth Uhealth 2019 Apr 22;7(4):e11831 [FREE Full text] [doi: 10.2196/11831] [Medline: 31008713]
41. 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]

Abbreviations

MARS: Mobile App Rating Scale
mHealth: mobile health

Edited by L Buis, G Eysenbach; submitted 09.09.20; peer-reviewed by D Saleh; comments to author 28.11.20; revised version received 12.12.20; accepted 12.02.21; published 04.03.21.

Please cite as:
Kabir MA, Rahman SS, Islam MM, Ahmed S, Laird C
Mobile Apps for Foot Measurement in Pedorthic Practice: Scoping Review
JMIR Mhealth Uhealth 2021;9(3):e24202
URL: https://mhealth.jmir.org/2021/3/e24202
doi:10.2196/24202
PMID:33661124

Original Paper

Comparing Smartphone Apps for Traditional Chinese Medicine and Modern Medicine in China: Systematic Search and Content Analysis

Xiao Hang Liu[1*], MD; Fan Jin[2*], MD; Jeffrey Hsu[1], MA, MD; Di Nan Li[3], BMed; Wei Chen[1], MD

[1]Department of Cardiology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

[2]Department of Infectious Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

[3]Yawlih Technology, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Wei Chen, MD
Department of Cardiology
Peking Union Medical College Hospital
Chinese Academy of Medical Sciences & Peking Union Medical College
No.1 Shuaifuyuan, East District
Beijing, 100730
China
Phone: 86 13716941630
Email: chenweipumch@126.com

Abstract

Background: Traditional Chinese medicine (TCM) is an integral part of mainstream medicine in China, with theories and practices that are completely different from modern medicine. TCM should not be ignored or confused with modern medicine in the analysis of the Chinese health care system, including the analysis of mobile health (mHealth) apps. To date, differences between TCM apps and modern medicine apps have not be systematically investigated.

Objective: The aim of this study was to systematically compare the quality of apps for TCM and modern medicine in China.

Methods: In December 2020, we searched iOS (iTunes) and Android (Tencent, Oppo, and Huawei app stores) platforms for all mHealth apps and then categorized them as TCM or modern medicine apps if they were included in the final analysis. The included apps were downloaded on smartphones and assessed by 2 reviewers on the following 4 aspects: (1) data in the app stores, including user ratings, download counts, cost, target users, and year of last update; (2) functionality; (3) quality of the app content as determined by the Mobile App Rating Scale (MARS); and (4) analysis of the app privacy and security.

Results: In total, 658 apps were analyzed, including 261 TCM medicine apps and 397 modern medicine apps. The average download count of modern medicine apps (approximately 5 million) was more than 10 times that of TCM apps (approximately 400,000). Regarding functionalities, 64.7% (257/397) of modern medicine apps provided telemedicine (74/261, 28.4% in TCM apps), 62.7% (249/397) provided registration (70/261, 26.8% in TCM apps), and 45.6% (181/397) provided communication (38/261, 14.6% in TCM apps). A larger proportion of TCM apps provided prescription and medication management (144/261, 55.2% in TCM apps versus 168/397, 42.3% in modern medicine apps). The majority of modern medicine apps (329/397, 82.9%) combined ≥3 functionalities compared with one-third of TCM apps (93/261, 34.6%). We then selected 81 top apps for quality and safety assessment (41 TCM apps and 40 modern medicine apps). Of these, the mean overall MARS score of TCM apps (2.7, SD 0.5) was significantly lower than modern medicine apps (3.6, SD 0.4). Almost all modern medicine apps (38/40, 95%) addressed privacy and security by providing a privacy policy and describing how to protect personal data, but less than half of the TCM apps (18/41, 44%) described this information (P<.001).

Conclusions: The different functionalities reflect the distinct innate characteristics of these two medical systems. Although great progress has been made and the Chinese mHealth market size is large, there still exist many opportunities for future development, especially for TCM.

(JMIR Mhealth Uhealth 2021;9(3):e27406) doi:10.2196/27406

KEYWORDS

mHealth; traditional Chinese medicine; modern medicine; mobile apps; app; comparison; content analysis; China; health care; development

Introduction

Traditional Chinese medicine (TCM) refers to a holistic medical system for the pathophysiology, diagnosis, treatment, and prevention of diseases [1]. As one of the oldest traditional medicine systems in the world, TCM was formed more than 2000 years ago and developed with the accumulation of knowledge and practice in the following centuries [2-5]. Outside China, TCM is treated as an important part of complementary and alternative medicine and has gained increasing attention [6]. In 2018, the World Health Organization (WHO) first included a chapter on TCM in the 11th revision of the International Classification of Diseases [7]. In China, with strong support from the government for the popularization of TCM and people's belief in traditional Chinese culture, there is no doubt that TCM is an integral part of mainstream medicine. According to the National Health Commission of the People's Republic of China, there were nearly 35,000 hospitals in 2019 in China, and 15.2% (5232) of them were hospitals of TCM. Meanwhile, there were 3267 outpatient departments and over 57,000 clinics of TCM. The TCM sector provided more than 1.1 billion medical services, accounting for 16.4% of health care in China [8]. During the COVID-19 pandemic, TCM also played an active role in the prevention of SARS-CoV-2, helped improve clinical symptoms of patients, and reduced the mortality rate [9].

The rapid development of information and communication technologies helps overcome the barrier of distance in the delivery of health care services, which enables the generation and proliferation of mobile health (mHealth) [10]. mHealth apps have epitomized typical mHealth service and have the potential to promote patient engagement, cut costs, and improve health outcomes [11]. In recent years, the establishment of telecommunication networks, continuous growth of smartphone usage rates, and increasing demand for high-quality and convenient health care services has led to significant development of the mHealth industry in China, including numerous and miscellaneous apps released. In 2020, the market size of mHealth reached ¥52.1 billion (approximately US $8.04 billion), and the number of mHealth users was 6.35 billion [12]. So far, several investigations providing an overview of the characteristics of mHealth apps in China have been published, but they did not distinguish whether the apps were tailored for TCM or modern medicine [13,14]. Given the unique medical system in China and natural differences in many aspects of TCM and modern medicine, including mechanisms of actions, mode of treatment, training of practitioners, quality of medicines, involvement of the healer and the patient, safety, and adverse effects [2], the performance and contribution of TCM should not be ignored or confused with modern medicine (also known as Western medicine in China) in the analysis of mHealth apps.

Thus, the aim of this study was to systematically compare the quality of apps for TCM and modern medicine in China. Specific objectives were to assess the basic characteristics, functionalities, app content (using the Mobile App Rating Scale [MARS]), and fairness of privacy policies. To our knowledge, no similar study has been done.

Methods

Search Criteria and App Selection

In December 2020, we conducted a thorough review of mHealth apps across the Apple iTunes app store (iOS). For Android, we sampled apps from the 3 largest Android app stores in China: Tencent Myapp, Oppo, and Huawei [15]. Unlike most other countries, the Google Play Store was blocked in China and the general public could not access Google Play and download apps. Therefore, we searched the most commonly used Chinese app stores for Android devices to reflect the real-world conditions in China. The following search terms were identified: "mobile health" OR "medicine" OR "traditional Chinese medicine" OR "traditional medicine" OR "modern medicine" OR "Western medicine." There were no restrictions concerning subcategories like "medical."

Screening was conducted based on app titles, marketing descriptions, and screenshots of the potential apps for relevance and inclusion. The following apps were excluded: (1) apps not relevant to our study purpose or designed for entertainment, product advertisement, loans, etc; (2) apps not in Chinese; (3) question banks or online guidance for examinations like the medical licensing examination, professional postexamination, and test for the national residency standard training program; (4) apps focused on hospital administration and management; and (5) apps pertaining to general health, for example, menstrual cycle management, sleep monitoring, and water intake reminder apps. The remaining apps were downloaded for eligibility. If an app could not run properly, it was also excluded. After selection, the included apps were classified as TCM apps or modern medicine apps based on the app's content. All apps were downloaded onto an iPhone 12 (version 14.2.1) and a Huawei Nova 2s (version 9.0). Two reviewers (XHL and FJ) performed the assessment of the apps using a standard data extraction form.

Assessment of Apps

First, the general characteristics of the apps were recorded in the database, including platform, average user-scored star rating, download counts, year of the last update, target user, and app cost. Then, we investigated the services each app provided through use of the app. The specific functionalities used in this study were telemedicine, registration, prescription and medication management, communication, records, citizen-based reporting, on-demand information services to clients, client financial transactions, decision support, health worker activity

planning and scheduling, health care provider training, and laboratory and diagnostics imaging management. The definition of these functionalities originated from the classification of digital health interventions proposed by the WHO and were tailored to this study [16,17].

For further analysis, we selected the top 25 Android apps and top 25 iOS apps from TCM and modern medicine apps, respectively. For Android, we chose by the number of downloads in descending order. For iOS, as information on the number of downloads was not provided, the order of selection was dependent on search retrieval order on the platform.

The MARS was then used to rate the selected apps' quality, including objective and subjective app quality evaluation [18]. The objective app quality section contained 19 evaluation criteria clustered within 4 domains: engagement, functionality, aesthetics, and information. The domain of subjective quality included 4 criteria to evaluate the overall satisfaction of users. Each evaluation criterion was scored on a 5-point Likert scale (1=unqualified, 2=poor, 3=acceptable, 4=good, 5=excellent). Two independent reviewers viewed the training video and tested each app for at least 10 minutes. After scoring all the evaluation criteria, the total mean MARS score describing the overall quality of the app was obtained by calculating the average value of the 5 domains.

Additionally, based on the guidance of privacy in mobile apps recommended by the Information Commissioner's Office [19] and the mobile app privacy and security best practices published by the Online Trust Alliance [20], the assessment of privacy and security consisted of 7 questions. By answering yes or no, the accessibility of the privacy policy and ability to protect personal data were assessed.

Any discrepancies in the assessment of apps were resolved by discussion with other researchers (DNL and JH) of the study team until consensus was reached.

Analysis of Apps Operated by Chinese Top Hospitals

The ranking list of Chinese hospitals is published every year to provide direction for discipline construction and guidance for patients [21,22]. Hospitals are rated on medical quality, resource allocation, academic research, and other criteria. In the apps fulfilling the inclusion criteria of our study, we first identified the number of apps operated by the top 100 TCM hospitals and top 100 modern medicine hospitals, respectively, according to the latest ranking list. Because appointment scheduling is the main function of apps directly operated by hospitals, we then investigated whether the patients needed to verify their identity before making an appointment and whether the apps provided appointment guidance, doctor selection, department selection, and online payment services during appointment scheduling.

Statistical Analysis

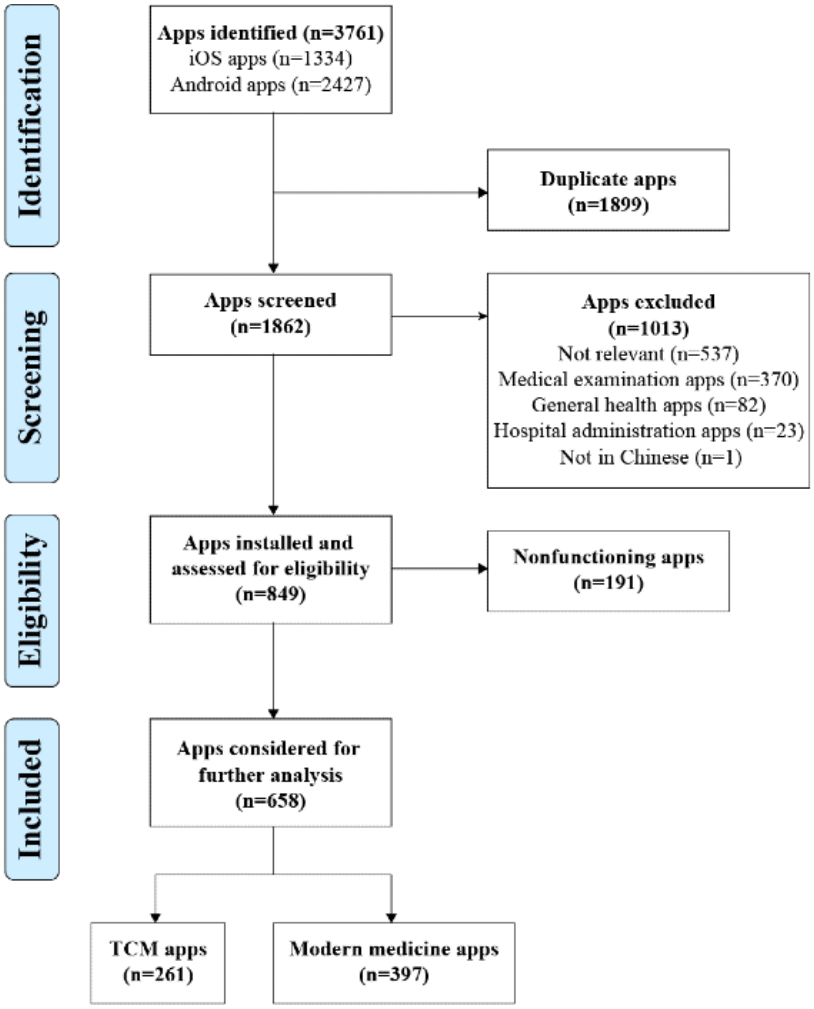
Categorical variables were compared using an uncorrected chi-square test or Fisher exact test. Continuous variables were analyzed using independent *t* tests. In the MARS evaluation, the Cohen test was performed to guarantee the reliability of the data analyzed by the 2 independent researchers. All statistical analyses were performed using IBM SPSS 22.0 (IBM Corp). A *P* value of <.05 was considered statistically significant.

Results

Summary of Search Results

iOS and Android app store searches identified 3761 potential apps, of which 1899 were removed as duplicates. Of the remaining 1862 apps, 658 met the indicated criteria. Among these included apps, 261 apps were classified as TCM apps and the other 397 focused specifically on modern medicine. The flow diagram (Figure 1) provides an overview of the selection process and reasons for exclusion. Major reasons for exclusion were that the app was not relevant to medicine (n=537) or it was a medical examination app (n=370).

Figure 1. Screening process flowchart. TCM: traditional Chinese medicine.



General Characteristics of Apps

Table 1 summarizes the general characteristics of the included TCM apps and modern medicine apps. There was no difference in overall user-scored rating between these two categories of apps (4.0, SD 1.1 vs 3.8, SD 1.2; *P*=.11). Although the overall user ratings were generally equal, modern medicine apps had a huge advantage in the number of downloads. In Tencent Myapp, the average number of downloads of modern medicine apps (approximately 5 million) was more than 10 times that of TCM apps (approximately 400,000). In addition, modern medicine apps were updated more frequently. A larger percentage of modern medicine apps were updated during the last year (352/397, 88.7% of modern medicine apps compared with 151/261, 57.9% of TCM apps), while 10.7% (28/261) of TCM apps had not been updated in the last 3 years. Most apps could be downloaded without cost, including 98.7% (392/397) of modern medicine and 91.2% (238/261) of TCM apps. There was a large difference in the constitution of target users. A total of 74.7% (195/261) of the TCM apps could be used by people who were not health care professionals (HCPs), but 41.6% (165/397) of the modern medicine apps were designed only for HCPs.

Table 1. General characteristics of TCM apps and modern medicine apps.

Characteristics	TCM[a] (n=261)	Modern medicine (n=397)	*P* value
Platform, n (%)			
iOS	109 (41.8)	94 (23.7)	<.001
Android	115 (44.1)	236 (59.4)	<.001
iOS and Android	37 (14.2)	67 (16.9)	.35
Ratings, mean (SD)			
iOS	4.0 (1.1)	4.5 (0.7)	<.001
Android	3.6 (1.3)	3.2 (1.3)	.03
iOS and Android[b]	4.4 (0.7)	4.5 (0.4)	.26
Overall	4.0 (1.1)	3.8 (1.2)	.11
Number of downloads[c] (thousand), n (%)			
>10,000	0 (0)	24 (7.9)	<.001
1000-9999	10 (6.6)	54 (17.8)	<.001
100-999	39 (25.7)	110 (36.3)	<.001
10-99	54 (35.5)	88 (29.0)	.65
<10	49 (32.2)	27 (8.9)	<.001
Year of last update, n (%)			
2020	151 (57.9)	352 (88.7)	<.001
2019	42 (16.1)	27 (6.8)	<.001
2018	18 (6.9)	9 (2.3)	.003
Before 2018	28 (10.7)	5 (1.3)	<.001
No updates	22 (8.4)	4 (1.0)	<.001
Cost, n (%)			
No	238 (91.2)	392 (98.7)	<.001
Yes	23 (8.8)	5 (1.3)	<.001

[a]TCM: traditional Chinese medicine.

[b]Ratings were calculated as the average on iOS and Android platforms.

[c]We only counted the number of downloads on Android because this was not applicable on iOS. TCM: n=152; modern medicine: n=303.

Functionalities of Apps

The detailed classification criteria of functionalities are shown in Figure 2. TCM and modern medicine apps displayed divergent functionalities (as shown in Table 2). As a whole, the most common service factors were telemedicine, identification and registration, and prescription and medication management. The least common service factors were laboratory and imaging management and activity planning and scheduling. By comparison, more modern medicine apps provided services such as registration, telemedicine, and communication. A larger proportion of TCM apps provided prescription and medication management. Specifically, not only were herbs and Chinese patent drugs available in the online pharmacy of TCM apps but the service of daily home delivery of decocted drugs was also provided by these apps because some herbs need to be boiled. Similar proportions of TCM and modern medicine apps provided on-demand information services for non-HCPs. Regarding the content of the information services, we noticed that many TCM apps provided relatively professional knowledge and self-screening. Professional knowledge of TCM was mainly conveyed by presenting and interpreting the classic TCM books, like *Inner Canon of the Yellow Emperor* (earliest medical classic in China) and *Compendium of Materia Medica* (an outline treatise of medical herbs). They provided e-books or audio files. Self-screening was usually done by answering questionnaires that showed the patients whether they had a yin deficiency or yang deficiency of certain organs. Using artificial intelligence, one app analyzed the health condition of a user's body after capturing and uploading their facial expression (Multimedia Appendix 1).

Figure 2. Classification criteria of functionalities. The criteria originated from the classification of digital health interventions proposed by the WHO and were tailored to this study. As opposed to the WHO criteria, prescription and medication management, identification and registration, and telemedicine were not under a single category of HCP or non-HCP because they could be mutual processes. Targeted client communication, untargeted client communication, client-to-client communication, and HCP communication were summarized as "communication." “Records” included personal health tracking and health records. Referral coordination was removed from our study because no app provided this service. HCP: health care professional; WHO: World Health Organization.

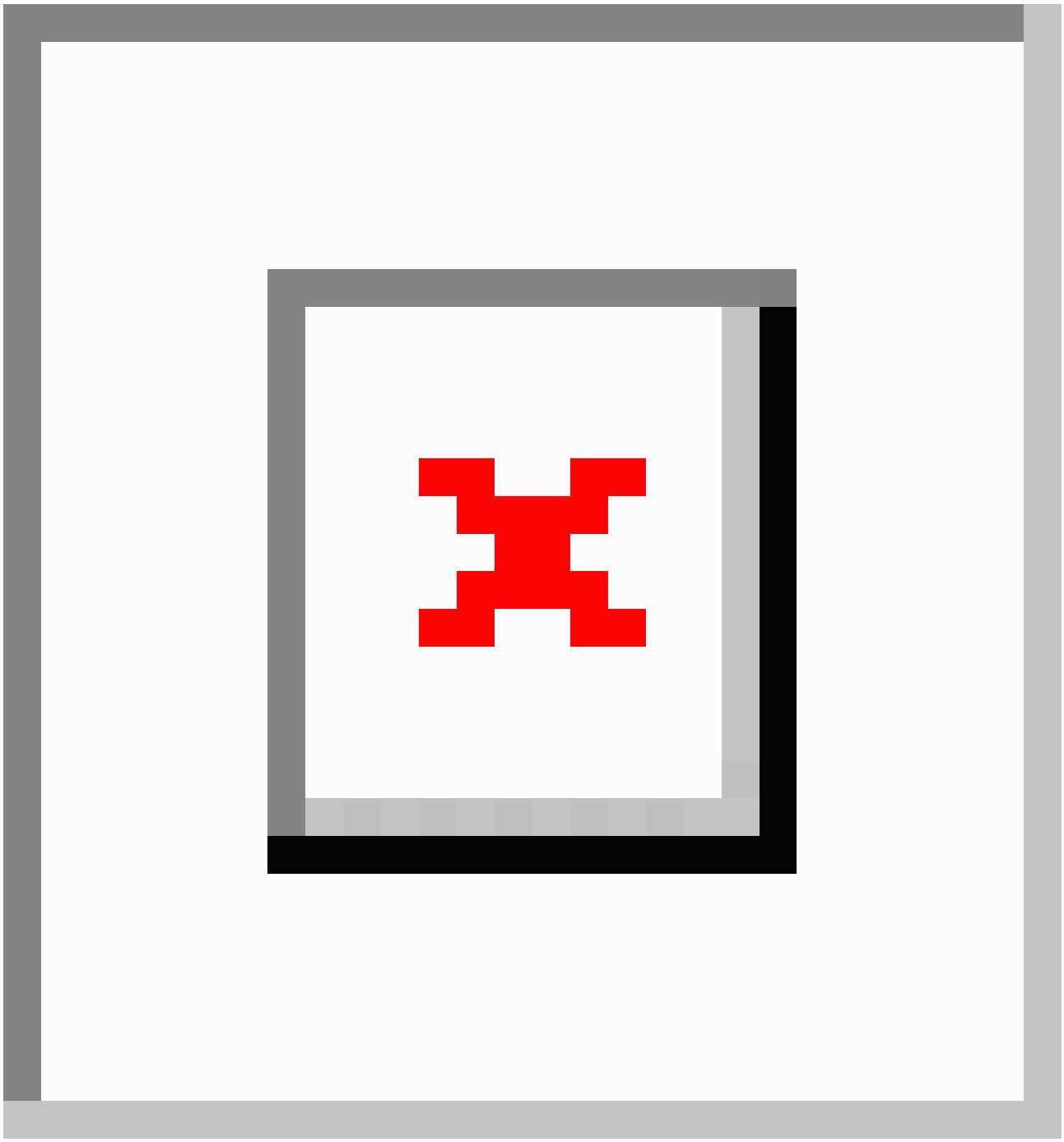


Table 2. Frequency of app functionalities.

Functionality	Total, n (%) (n=658)	TCM[a], n (%) (n=261)	Modern medicine, n (%) (n=397)	*P* value
Prescription and medication management	312 (47.4)	144 (55.2)	168 (42.3)	.001
Identification and registration	319 (48.4)	70 (26.8)	249 (62.7)	<.001
Telemedicine	331 (50.3)	74 (28.4)	257 (64.7)	<.001
Communication	219 (33.3)	38 (14.6)	181 (45.6)	<.001
Records	149 (22.6)	24 (9.2)	125 (31.5)	<.001
Citizen-based reporting	160 (24.3)	56 (21.5)	104 (26.2)	.16
On-demand information services	286 (43.4)	106 (40.6)	180 (45.3)	.23
Financial transactions	210 (31.9)	62 (23.8)	148 (37.3)	<.001
Decision support	164 (24.9)	43 (16.5)	121 (30.5)	<.001
Activity planning and scheduling	54 (8.2)	12 (4.5)	42 (10.5)	.006
Training	130 (19.8)	52 (19.9)	78 (19.6)	.93
Laboratory and imaging management	46 (6.9)	10 (3.8)	36 (9.0)	.01

[a]TCM: traditional Chinese medicine.

We further analyzed the combinations of app functionalities. Figures 3 and 4 show all the combinations of TCM apps and modern medicine apps. Compared with TCM apps, modern medicine apps provided more comprehensive functionalities and various combinations. The majority of modern medicine apps (329/397, 82.9%) combined ≥3 functionalities compared with about one-third of TCM apps (93/261, 35.6%). Modern medicine apps also produced more combination patterns than TCM apps (39 vs 32). The most common combinations of functionalities in TCM apps were (1) prescription and medication management plus training and (2) prescription and medication management plus on-demand information services, while the most common combinations of functionalities in modern medicine apps were (1) decision support plus training and (2) telemedicine plus identification and registration plus records.

Figure 3. TCM apps' functionalities. The pie chart shows the percentage of TCM apps' target users. The stacked bar chart shows the distribution of TCM apps according to key functionalities and user type. The upset plot shows the intersection of multiple functionalities. TCM: traditional Chinese medicine.

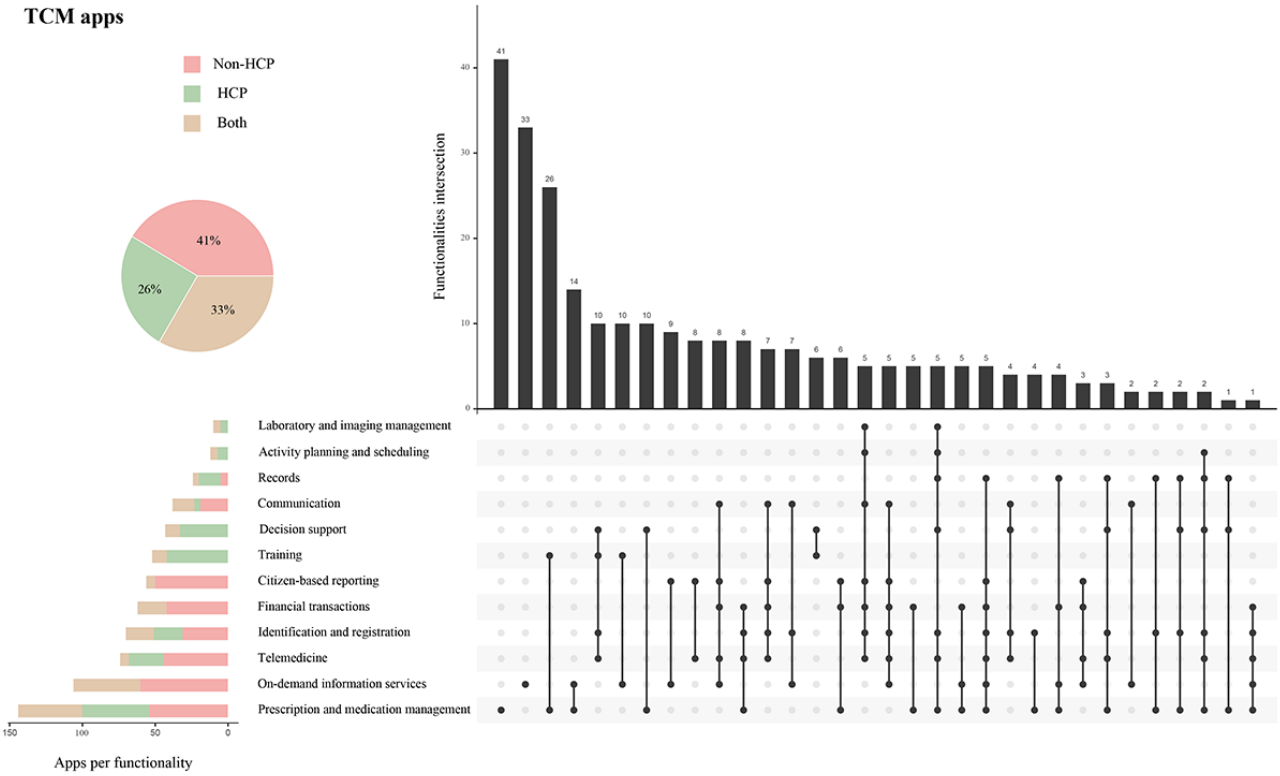
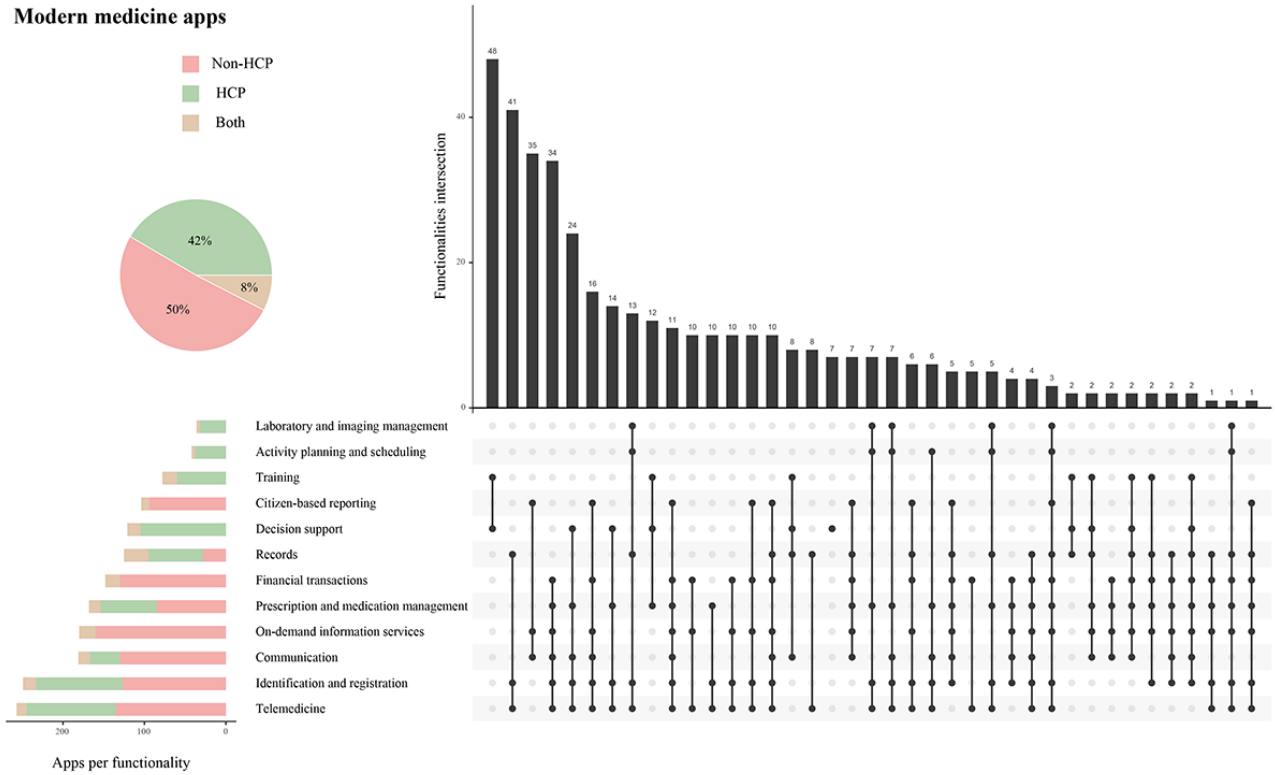


Figure 4. Modern medicine apps' functionalities. Pie chart: portion of modern medicine apps' target users. Stacked bar chart: distribution of modern medicine apps according to key functionalities and user type. Upset plot: intersection of multiple functionalities.



MARS Evaluation of Top Apps

Among the top 50 apps of the two categories, 9 TCM apps and 10 modern medicine apps were available on both iOS and Android platforms. Therefore, 41 TCM apps and 40 modern medicine apps were included in the MARS evaluation.

The mean overall MARS score for the TCM apps was significantly lower than that for the modern medicine apps (2.7, SD 0.5 vs 3.6, SD 0.4; *P*<.001) (Table 3). The percentage of apps that scored higher than the minimum acceptability score of 3.0 was 34% (14/41) of TCM apps and 95% (38/40) of modern medicine apps. A total of 5 TCM apps scored ≤2 points and none scored ≥4 points. By contrast, among the modern medicine apps, no apps scored ≤2 points and 4 scored ≥4 points.

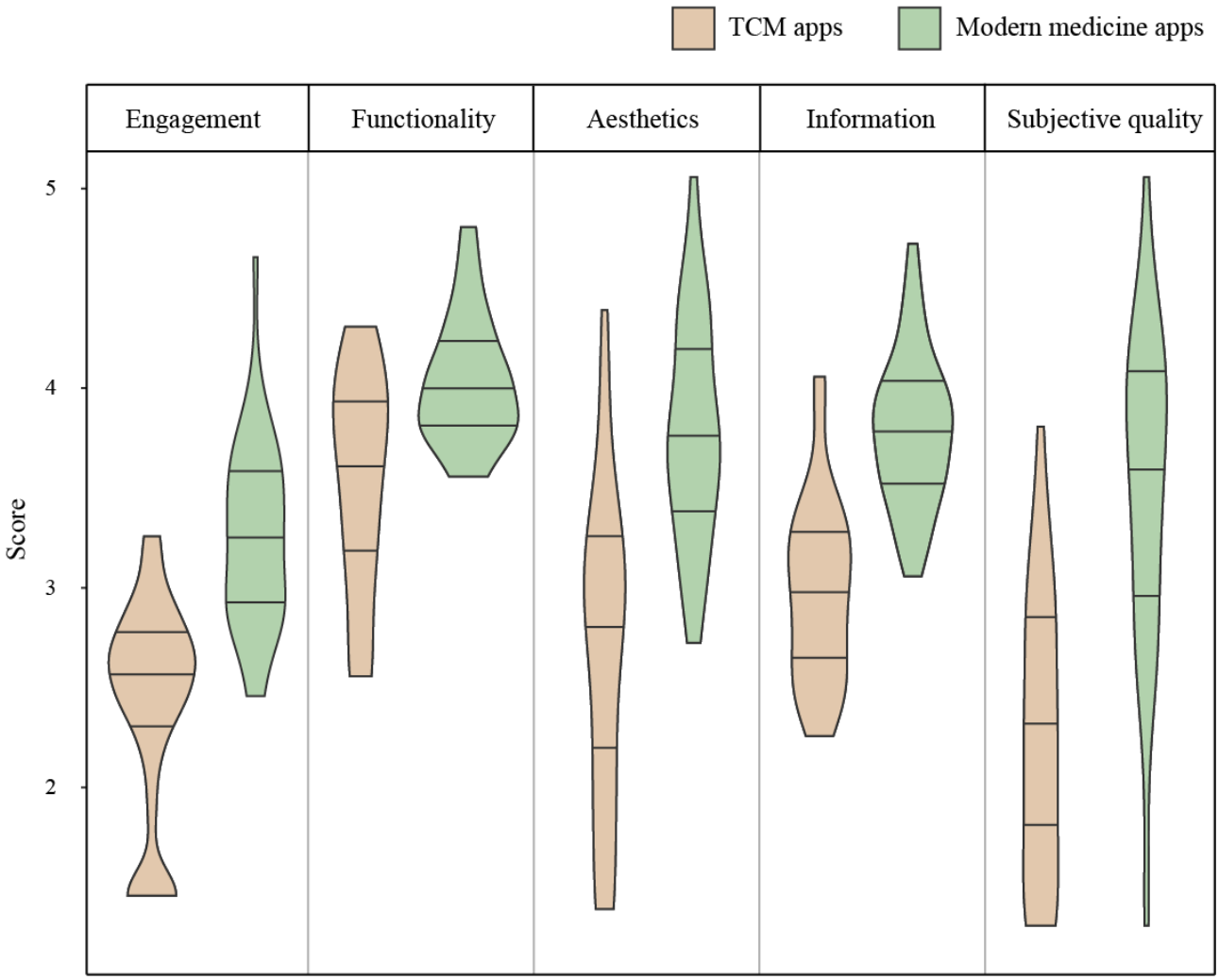
Table 3. Comparison of TCM and modern medicine apps' results for the Mobile App Rating Scale evaluation.

Category	TCM[a], mean (SD) (n=41)	Modern medicine, mean (SD) (n=40)	*P* value
Engagement	2.3 (0.5)	3.2 (0.5)	<.001
Functionality	3.5 (0.6)	4.0 (0.3)	<.001
Aesthetics	2.6 (0.8)	3.8 (0.5)	<.001
Information	2.9 (0.5)	3.7 (0.4)	<.001
Subjective quality	2.2 (0.7)	3.4 (0.8)	<.001
Overall	2.7 (0.5)	3.6 (0.4)	<.001

[a]TCM: traditional Chinese medicine.

In all 5 domains (engagement, functionality, aesthetics, information, and subjective quality), modern medicine apps outperformed TCM apps, and the differences were statistically significant (*P*<.001). Functionality had the highest mean score in both TCM apps (3.5, SD 0.6) and modern medicine apps (4.0, SD 0.3), while engagement was the poorest in the objective quality for both kinds of apps (3.2, SD 0.5 in modern medicine; 2.3, SD 0.5 in TCM). The Cohen coefficient was 0.92, indicating excellent interrater reliability. Figure 5 presents the specific score distribution of the two categories of apps in each domain.

Figure 5. Violin plot of the Mobile Application Rating Scale section item scores. The 3 black lines in each plot present the first quartile, the median, and the third quartile (from bottom to top). TCM: traditional Chinese medicine.



Privacy and Security of Top Apps

The privacy policy was available for 68% (28/41) of TCM apps and 95% (38/40) of modern medicine apps (P=.02). In comparison with TCM apps, more modern medicine apps (17/40, 43% vs 6/41, 15%; P=.008) had a short-form privacy and security notice that highlighted key data practices. More modern medicine apps collected personally identifiable information and shared the data with third parties. Additionally, modern medicine apps performed better in data safety protection. A total of 95% (38/40) of the modern medicine apps described how the personal data were protected, but only 44% (18/41) of TCM apps described this information (P<.001) (Table 4).

Table 4. Assessment of privacy and security in regard to data gathering, sharing, and security as described in the privacy policy.

Privacy and security question	Total, n (%) (n=81)	TCM[a], n (%) (n=41)	Modern medicine, n (%) (n=40)	*P* value
Is the privacy policy available without the need to download the app?				.02
No	15 (18.5)	13 (31.7)	2 (5.0)	
Yes	66 (81.5)	28 (68.3)	38 (95.0)	
Is the privacy policy available within the app?				.004
No	24 (29.6)	18 (43.9)	6 (15.0)	
Yes	57 (70.4)	23 (56.1)	34 (85.0)	
Is there a short-form notice highlighting key data practices?				.008
No	48 (59.3)	27 (65.9)	21 (52.5)	
Yes	23 (28.4)	6 (14.6)	17 (42.5)	
Not applicable	10 (12.3)	8 (19.5)	2 (5.0)	
Is the privacy policy available in any other language?				.23
No	78 (96.3)	41 (100)	37 (92.5)	
Yes	3 (3.7)	0 (0)	3 (7.5)	
Does the app collect personally identifiable information?				.004
No	14 (17.3)	12 (29.3)	2 (5.0)	
Yes	65 (80.2)	27 (65.9)	38 (95.0)	
Not specified	2 (2.5)	2 (4.9)	0 (0)	
Does the app share users' data with a third party?				<.001
No	16 (19.8)	11 (26.8)	5 (12.5)	
Yes	44 (54.3)	12 (29.3)	32 (80.0)	
Not specified	21 (25.9)	18 (43.9)	3 (7.5)	
Does the app say how the users' data security is ensured (eg, encryption, authentication, fire wall)?				<.001
No	25 (30.9)	23 (56.1)	2 (5.0)	
Yes	56 (69.1)	18 (43.9)	38 (95.0)	

[a]TCM: traditional Chinese medicine.

Comparison of Hospital Apps

Among the top hospitals in China, 41 hospitals of modern medicine had their own apps, but the number sharply decreased to 11 for hospitals of TCM. As expected, all these apps provided the service of registration. Although significantly fewer hospitals of TCM developed apps for patients, regarding the procedure of appointment scheduling, all hospital apps, no matter if they were TCM or modern medicine apps, provided the services of department selection and doctor selection so that the patients could visit specific doctors according to the patients' preferences and the doctors' areas of expertise. Identity verification was also required in all apps via identity card or phone number. This could help stop scalpers making profits from scheduling fake appointments. Due to the rapid development of e-commerce in China, all TCM apps and 33 of 41 (81%) modern medicine apps allowed online payment via Alipay, Wechat, and debit card. A total of 78% (32/41) of modern medicine apps and 64% (7/11) of TCM apps provided appointment guidance, but the difference was not statistically significant (*P*=.56).

Discussion

Principal Findings

This study systematically compared the apps designed for TCM and modern medicine available in China. Considering that TCM is a traditional medicine system and mHealth is a newly developed information technology and taking into account the important role of TCM in Chinese culture and the Chinese medical system, TCM should not be overlooked or simply confused with modern medicine in the analysis of mHealth apps. Overall, our findings suggest that there are currently a considerable number of both TCM (n=261) and modern medicine (n=397) apps on the market, but TCM apps and modern medicine apps had distinct functionalities and

combinations of functionalities. The TCM apps scored lower in all aspects (engagement, functionality, aesthetics, information quality, and subjective quality) of the multidimensional measure of app quality using the MARS app rating tool. Great progress has been made regarding the privacy and security of mHealth apps in China. We discuss our findings further below.

The Android app stores with the largest number of monthly active users in China are operated by Tencent, Oppo, and Huawei. Google Play is difficult to access and has been almost absent since 2010 [13,23]. Thus, we sampled apps from Tencent Myapp, Oppo App Store, Huawei App Store, and Apple iTunes App Store and applied keywords commonly used for TCM and modern medicine to ensure the representativeness of the apps we included. Our study displayed a current landscape of 261 TCM apps and 397 modern medicine apps. More than half of the modern medicine apps had been updated in the last year to improve performance, fix bugs, and update information to adapt to rapidly growing medical knowledge. TCM apps and modern medicine apps had similar overall ratings, but download counts of modern medicine apps were much higher. Nearly 10% of modern medicine apps had over 10 million downloads, and the top app, named Ping An Good Doctor, had more than 200 million downloads. In contrast, more than half of TCM apps did not reach the download volume of 100,000. This result indicates that although a certain number of TCM apps have been developed, the subsequent promotion, usage, and software maintenance remain problems. The majority of people are used to the traditional offline modality of TCM rather than the use of mHealth apps.

As displayed in the functionalities and combinations of functionalities, both TCM and modern medicine apps attached great importance to on-demand information services to clients. The differences in functionalities reflected the differences in medical theories and practices between TCM and modern medicine. The mechanism of modern medicine is based on anatomy, pathophysiology, molecular biology, and other basic medical and clinical medical knowledge, while the formation of TCM is on the basis of ancient Chinese philosophy. The diagnosis of a certain disease in Western medicine needs modern equipment and laboratory testing, while diagnosing a *zheng* in TCM relies more on the doctor's experience and observation without drawing blood or doing radiological examinations [24]. A *zheng* (syndrome) is an outcome after analyzing all symptoms and signs. One disease in modern medicine may have several *zheng* and a *zheng* could be caused by different diseases [25]. Modern medicine primarily treats patients through medicine or surgery with additional information about precautions and side effects, while TCM treatment approaches include herbs, minerals, and guidelines on lifestyle [26]. Therefore, preliminary understanding of TCM is easier for the public, and the gap between HCPs and non-HCPs is smaller. As a result, modern medicine apps pay more attention to building connections between doctors and patients, while TCM apps tend to educate the patients more. In our study, 41.6% (165/397) of modern medicine apps were designed for HCPs, while 74.7% (195/261) of TCM apps' target users included non-HCPs.

Similarly, more modern medicine apps provided telemedicine and appointment-making services. In China, patients can purchase Chinese patent drugs without strict restrictions, which is much easier than buying prescription drugs. This explains why medication management appeared more commonly in TCM apps. Because laboratory tests and imaging examinations are important parts of modern medicine, it was no surprise that laboratory and diagnostics imaging management services were more common in modern medicine apps.

We extracted top TCM and modern medicine apps for further analysis of quality and security. MARS is a systematic and validated questionnaire that is not too technical or specific to a particular health domain [18,27]. Our study showed a mean score of 2.7 and 3.6 for the overall quality of TCM apps and modern medicine apps, respectively. A rating of ≥3 points indicates overall acceptable quality [28]. All top modern medicine apps scored a value of ≥3 points, and 4 of 41 (10%) apps exceeded 4 points. However, half of TCM apps did not reach the acceptable level. The score of Western medicine apps was similar or even higher compared with the results of other studies. Davalbhakta et al [29] showed a rating of 3.7 in apps for the management of COVID-19. Kim et al [30] found the mean MARS score for the overall quality of apps to be 3.23 for potential drug-drug interaction checks. TCM apps scored much lower in both objective and subjective quality assessments. Differences between TCM apps and modern medicine apps were statistically significant for all dimensions of the MARS. The engagement domain scored poorest in both categories, which is consistent with previous studies of apps targeting genitourinary tumors and COVID-19 [29,31]. This could be explained by the primary purpose of mHealth apps. Because the main target users of mHealth apps are either doctors or patients, it is difficult to make them feel interested and relaxed when facing diseases. Apps focusing on behavior change might score higher on the engagement dimension because the engagement of users is a key factor of successful behavior change [32].

Privacy policies and data security are particularly important for mHealth apps because they usually collect personally identifiable information and data related to the users' health conditions. It is disappointing that privacy policies were absent within and outside of the app (on the website or app store) in 9 of the top 41 (22%) TCM apps and that 23 of the 41 (56.1%) apps did not mention how they ensured the users' data security at all. However, we believe this is a positive fact for two reasons. First, according to Hsu et al [13], in December 2015, nearly all Chinese top mHealth apps, let alone the remaining apps, lacked information security. The Chinese government did not have a direct policy or documents on mHealth security then [13]. In June 2016, the general office of the People's Republic of China State Council issued guidance on promoting and standardizing the development of health care applications. One of the major principles was keeping a balance between app development and its safety and protecting individual privacy and information security effectively [33]. It is encouraging that great progress has been made in recent years, starting from scratch. Second, we included various types of apps in our study. The main function of 6 of the 9 apps without privacy policies was health care information provision to clients, and they did not collect personally identifiable information. Likewise, Sunyaev et al

[34] found a surprising result in that only 31% of medical or health and fitness apps had privacy policies. Privacy policies might be less frequently provided in apps handling less sensitive information about patients. In terms of modern medicine apps, all of them incorporated or linked to a privacy policy, which is even better than the findings of Huckvale et al [35], who studied top-ranked apps for depression and smoking cessation on Android and iOS platforms. They assessed privacy policies in a more detailed way according to a schema of privacy policy quality criteria [36]. In this study, we managed to provide a comprehensive comparison between TCM apps and modern medicine apps instead of deeply investigating their privacy policies.

Future Perspectives

There exist many opportunities for further development of both TCM and modern medicine apps in China in the future. On the one hand, for TCM, mHealth is a perfect tool that has not been fully used. The basic diagnostic procedure of TCM is composed of 4 techniques: looking, listening and smelling, questioning, and feeling the pulse. The first 3 methods can be easily realized using photos and videos via remote communication technologies on mHealth apps. Moreover, Tang et al [37] reported an electronic TCM pulse diagnostic system developed with an artificial neural network. Considering that herbs and Chinese patent drugs are not strictly restricted and there is a well-developed express industry in China, diagnosing and treating without seeing patients face to face is more realizable for TCM. TCM hospitals should have made more efforts to develop online services, but in our study, only 11 of the top 100 TCM hospitals had their own apps in service. In addition, in order to make TCM information more evidence based and accepted both domestically and internationally, the Chinese government and TCM experts should keep investing in programs devoted to the modernization and standardization of TCM. On the other hand, for modern medicine, the online-to-offline approach (an integration of offline businesses into online commerce) has been relatively mature [38]. In the future, internet hospitals should be further developed to break the reliance on traditional health care providers [39].

Limitations

There are limitations to this study. First, we did not access Google Play for the Android apps, which could cause selection bias. Google Play is widely used worldwide. However, the general public in mainland China could not download or purchase apps on Google Play at all. Therefore, we selected the 3 largest Android app stores in China to reflect the real-world conditions [15]. Second, sometimes it was difficult to put an app into the category of TCM or modern medicine because the apps (less than 10 in our study) provided both kinds of services simultaneously. We fully assessed all the functions and discussed to determine the main focus of these apps. If disagreement existed, a third experienced investigator was invited. Third, we excluded general health apps to achieve a more homogeneous analysis because we could not define whether general health apps belonged to TCM or modern medicine. Finally, under the category of both TCM and modern medicine, we presented broad coverage and apps for various initiatives were included. It is worth mentioning that the main purpose of this study was to provide a global analysis of the distribution of and differences between TCM apps and modern medicine apps rather than apps for a single disease or service.

Conclusions

We identified the number and functions and evaluated the quality and privacy of TCM apps and modern medicine apps currently available on the Chinese market. mHealth in China is already a large market for both TCM and modern medicine, but there is still great potential for development. Different functionalities reflected the distinct innate characteristics of these two medical systems. Apps for modern medicine outperformed TCM apps in all aspects of quality assessment using MARS. TCM apps need quality improvement for further penetration into the market. It is gratifying that the developers, data controllers, and government paid attention to the privacy and security of mHealth apps. This work can inform the future development of mHealth apps for developers and provide an important reference for researchers and customers to search, review, and compare the TCM and modern medicine apps in China.

Acknowledgments

This study was supported by a grant from Capita's Funds for Health Improvement and Research (CFH 2020-2-40110).

Authors' Contributions

XHL and JH were responsible for study conceptualization. XHL and FJ wrote the draft manuscript. XHL, FJ, and DNL rated apps. FJ did the statistical analysis and visualization. WC revised the manuscript. All authors reviewed the draft and provided comments for changes. All authors approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshot of a TCM app for self-screening.
[PNG File , 155 KB - mhealth_v9i3e27406_app1.png]

References

1. Xu Q, Bauer R, Hendry BM, Fan T, Zhao Z, Duez P, et al. The quest for modernisation of traditional Chinese medicine. BMC Complement Altern Med 2013 Jun 13;13:132 [FREE Full text] [doi: 10.1186/1472-6882-13-132] [Medline: 23763836]
2. Raja Ikram RR, Abd Ghani MK, Abdullah N. An analysis of application of health informatics in Traditional Medicine: A review of four Traditional Medicine Systems. Int J Med Inform 2015 Nov;84(11):988-996. [doi: 10.1016/j.ijmedinf.2015.05.007] [Medline: 26160148]
3. Tang J, Liu B, Ma K. Traditional Chinese medicine. Lancet 2008 Dec;372(9654):1938-1940. [doi: 10.1016/s0140-6736(08)61354-9]
4. Lu A, Jia HW, Xiao C, Lu QP. Theory of traditional Chinese medicine and therapeutic method of diseases. World J Gastroenterol 2004 Jul 01;10(13):1854-1856 [FREE Full text] [doi: 10.3748/wjg.v10.i13.1854] [Medline: 15222022]
5. Wang J, Wong Y, Liao F. What has traditional Chinese medicine delivered for modern medicine? Expert Rev Mol Med 2018 May 11;20:e4. [doi: 10.1017/erm.2018.3] [Medline: 29747718]
6. Wieland LS, Manheimer E, Berman BM. Development and classification of an operational definition of complementary and alternative medicine for the Cochrane collaboration. Altern Ther Health Med 2011;17(2):50-59 [FREE Full text] [Medline: 21717826]
7. ICD-11: Classifying disease to map the way we live and die. World Health Organization. 2018. URL: https://www.who.int/zh/news-room/spotlight/international-classification-of-diseases [accessed 2020-12-28]
8. Statistical bulletin on healthcare development. National Health Commission of the People's Republic of China. 2020. URL: https://www.nhc.gov.cn/guihuaxxs/s10748/202006/ebfe31f24cc145b198dd730603ec4442.shtml [accessed 2020-12-28]
9. Li Q, Wang H, Li X, Zheng Y, Wei Y, Zhang P, et al. The role played by traditional Chinese medicine in preventing and treating COVID-19 in China. Front Med 2020 Oct;14(5):681-688 [FREE Full text] [doi: 10.1007/s11684-020-0801-x] [Medline: 32651936]
10. Scott Kruse C, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: A systematic review. J Telemed Telecare 2018 Jan;24(1):4-12 [FREE Full text] [doi: 10.1177/1357633X16674087] [Medline: 29320966]
11. Sama PR, Eapen ZJ, Weinfurt KP, Shah BR, Schulman KA. An evaluation of mobile health application tools. JMIR Mhealth Uhealth 2014 May;2(2):e19 [FREE Full text] [doi: 10.2196/mhealth.3088] [Medline: 25099179]
12. Research report on the development of Chinese mHealth industry in 2020. iiMedia Research. 2020. URL: https://www.iimedia.cn/c400/75770.html [accessed 2021-01-17]
13. Hsu J, Liu D, Yu YM, Zhao HT, Chen ZR, Li J, et al. The Top Chinese Mobile Health Apps: A Systematic Investigation. J Med Internet Res 2016 Aug 29;18(8):e222 [FREE Full text] [doi: 10.2196/jmir.5955] [Medline: 27573724]
14. Lv Q, Jiang Y, Qi J, Zhang Y, Zhang X, Fang L, et al. Using Mobile Apps for Health Management: A New Health Care Mode in China. JMIR Mhealth Uhealth 2019 Jun 03;7(6):e10299 [FREE Full text] [doi: 10.2196/10299] [Medline: 31162131]
15. Chinese App Store List: Top 10 App Stores in China. 24hChina. 2019. URL: https://www.24hchina.com/chinese-app-store-list/ [accessed 2019-03-12]
16. Classification of digital health interventions v1.0. World Health Organization. 2018. URL: https://apps.who.int/iris/bitstream/handle/10665/260480/WHO-RHR-18.06-eng.pdf;jsessionid=4E2A2DC3A5039E6D9051BC5ABBF6B873?sequence=1 [accessed 2021-02-28]
17. Paiva JOV, Andrade RMC, de Oliveira PAM, Duarte P, Santos IS, Evangelista ALDP, et al. Mobile applications for elderly healthcare: A systematic mapping. PLoS One 2020;15(7):e0236091 [FREE Full text] [doi: 10.1371/journal.pone.0236091] [Medline: 32730266]
18. 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;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
19. Privacy in mobile apps: Guidance for app developers. Information Commissioner's Office. URL: https://ico.org.uk/media/for-organisations/documents/1596/privacy-in-mobile-apps-dp-guidance.pdf [accessed 2019-05-10]
20. Online Trust Alliance. 2014. URL: https://www.internetsociety.org/resources/ota/2017/mobile-app-privacy-security/ [accessed 2021-03-20]
21. Chinese top 100 TCM hospitals in 2019. GZ Asclepius Healthcare. URL: http://www.ailibi.com/web/rank [accessed 2020-08-31]
22. Chinese Hospital Ranking in 2019. Fudan University Hospital Management Institute. URL: http://www.fudanmed.com/institute/news2019-2.aspx [accessed 2020-11-14]
23. Matthews C. Google is Heading Back to China. Time Magazine. 2015 Sep 04. URL: http://time.com/4023367/google-china/ [accessed 2020-12-28]
24. Gu P, Chen H. Modern bioinformatics meets traditional Chinese medicine. Brief Bioinform 2014 Nov;15(6):984-1003. [doi: 10.1093/bib/bbt063] [Medline: 24067932]
25. Wang Z, Li J, Xie Y, Yu X, Zhang Z. Traditional Chinese medicine ZHENG identification of bronchial asthma: Clinical investigation of 2500 adult cases. Complement Ther Med 2017 Feb;30:93-101. [doi: 10.1016/j.ctim.2016.12.003] [Medline: 28137533]

26. Telles S, Pathak S, Singh N, Balkrishna A. Research on traditional medicine: what has been done, the difficulties, and possible solutions. Evid Based Complement Alternat Med 2014;2014:495635 [FREE Full text] [doi: 10.1155/2014/495635] [Medline: 25013445]
27. Steeb T, Wessely A, French L, Heppt M, Berking C. Skin Cancer Smartphone Applications for German-speaking Patients: Review and Content Analysis Using the Mobile App Rating Scale. Acta Derm Venereol 2019 Oct 01;99(11):1043-1044 [FREE Full text] [doi: 10.2340/00015555-3240] [Medline: 31233185]
28. 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]
29. 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]
30. 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]
31. Amor-García M, Collado-Borrell R, Escudero-Vilaplana V, Melgarejo-Ortuño A, Herranz-Alonso A, Arranz Arija J, 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]
32. 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]
33. Guidance on promoting and standardizing the development of health care applications. People's Republic of China State Council. 2016. URL: https://www.gov.cn/zhengce/content/2016-06/24/content_5085091.htm [accessed 2020-12-28]
34. Sunyaev A, Dehling T, Taylor PL, Mandl KD. Availability and quality of mobile health app privacy policies. J Am Med Inform Assoc 2015 Apr;22(e1):e28-e33 [FREE Full text] [doi: 10.1136/amiajnl-2013-002605] [Medline: 25147247]
35. Huckvale K, Torous J, Larsen ME. Assessment of the Data Sharing and Privacy Practices of Smartphone Apps for Depression and Smoking Cessation. JAMA Netw Open 2019 Apr 05;2(4):e192542 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.2542] [Medline: 31002321]
36. Huckvale K, Prieto JT, Tilney M, Benghozi P, Car J. Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. BMC Med 2015;13:214 [FREE Full text] [doi: 10.1186/s12916-015-0444-y] [Medline: 26404673]
37. Tang AC, Chung JW, Wong TK. Digitalizing traditional chinese medicine pulse diagnosis with artificial neural network. Telemed J E Health 2012;18(6):446-453. [doi: 10.1089/tmj.2011.0204] [Medline: 22732024]
38. Zhang J. Customer loyalty forming mechanism of O2O e-commerce. Int J Bus Soc Sci 2014;5(1):164-169 [FREE Full text]
39. Xie X, Zhou W, Lin L, Fan S, Lin F, Wang L, et al. Internet Hospitals in China: Cross-Sectional Survey. J Med Internet Res 2017 Jul 04;19(7):e239 [FREE Full text] [doi: 10.2196/jmir.7854] [Medline: 28676472]

Abbreviations

HCP: health care professional
MARS: Mobile App Rating Scale
mHealth: mobile health
TCM: traditional Chinese medicine
WHO: World Health Organization

Edited by G Eysenbach; submitted 25.01.21; peer-reviewed by J Zhang; comments to author 16.02.21; revised version received 10.03.21; accepted 15.03.21; published 24.03.21.

Please cite as:
Liu XH, Jin F, Hsu J, Li DN, Chen W
Comparing Smartphone Apps for Traditional Chinese Medicine and Modern Medicine in China: Systematic Search and Content Analysis
JMIR Mhealth Uhealth 2021;9(3):e27406
URL: https://mhealth.jmir.org/2021/3/e27406
doi:10.2196/27406
PMID:33759786

Original Paper

A Social Media–Promoted Educational Community of Joint Replacement Patients Using the WeChat App: Survey Study

Xianzuo Zhang[1], MD; Xiaoxuan Chen[2], PhD; Nikolaos Kourkoumelis[3], PhD; Ran Gao[4], MSc; Guoyuan Li[1], MD; Chen Zhu[1], MD, PhD

[1]Department of Orthopedics, The First Affiliated Hospital, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China

[2]College of Chemistry and Chemical Engineering, Xiamen University, Xiamen, China

[3]Department of Medical Physics, School of Health Sciences, University of Ioannina, Ioannina, Greece

[4]Department of Applied Psychology, School of Humanities, Guangdong Peizheng College, Guangzhou, China

Corresponding Author:
Chen Zhu, MD, PhD
Department of Orthopedics
The First Affiliated Hospital, Division of Life Sciences and Medicine
University of Science and Technology of China
Lujiang Road 17#
Luyang District
Hefei, 230022
China
Phone: 86 55162283338
Email: zhuchena@ustc.edu.cn

Abstract

Background: Much effort has been made to optimize the results of total hip arthroplasty and total knee arthroplasty. With the rapid growth of social media use, mobile apps, such as WeChat, have been considered for improving outcomes and patient satisfaction after total hip arthroplasty and total knee arthroplasty.

Objective: We aimed to evaluate the effectiveness of a WeChat-based community as an intervention for overall patient satisfaction.

Methods: The study was conducted among discharged in-hospital patients who received hip or knee procedures in the First Affiliated Hospital of the University of Science and Technology of China from April 2019 to January 2020. An educational online social community was constructed with the WeChat app. Participants willing to join the community were enrolled in a WeChat group and received 3 months of intervention and follow-up. Those who were not willing to use the account were included in a control group and received routine publicity via telephone, mail, and brochures. The Danish Health and Medicine Authority patient satisfaction questionnaire was used to score perioperative patient education and overall satisfaction. The contents in the group chat were analyzed using natural language processing tools.

Results: A total of 3428 patients were enrolled in the study, including 2292 in the WeChat group and 1236 in the control group. Participants in the WeChat group had higher overall satisfaction scores than those in the control group (mean 8.48, SD 1.12 vs mean 6.66, SD 1.80, P<.001). The difference between the two groups was significant for primary surgery based on subgroup stratification. To control confounding factors and explore the effects of WeChat participation as a mediating variable between perioperative patient education and overall satisfaction, hierarchical regression was utilized. An interpatient interaction model was found in the community group chat, and it contributed to overall satisfaction. Patients in the group with more interpatient interactions were more likely to have better overall satisfaction.

Conclusions: The social media–promoted educational community using WeChat was effective among joint replacement patients. Provision of more perioperative education is associated with more active patient participation in the community and therefore more patient satisfaction in terms of the overall joint procedure. Community group chat could facilitate interactions among patients and contribute to overall satisfaction.

(JMIR Mhealth Uhealth 2021;9(3):e18763) doi:10.2196/18763

KEYWORDS

WeChat; social media; arthroplasty; perioperative education; patient satisfaction

Introduction

Joint diseases of the lower limb extremities influence the quality of life of elderly people and are becoming increasingly frequent owing to the rising average life expectancy. To date, total hip arthroplasty and total knee arthroplasty have been the most successful surgical treatments for end-stage arthritis. Considerable progress has been made to optimize the end-to-end process by improving the surgical technique, rehabilitation, and perioperative care. Although the length of hospital stay has been shortened in recent years [1], rehabilitation remains a long-term procedure and a physiopsychological challenge for patients who have undergone arthroplasty. Continuous rehabilitation guidance has proved its efficacy during long-term follow-up [2-4]. However, most of the rehabilitation advice in China is given during hospital stay rather than in a community center. Since patients often have limited awareness and poor compliance after discharge, educational follow-up needs to be strengthened further.

Perioperative education increases patient satisfaction [5,6], but routine guidance via telephone and mail can be laborious and challenging for elderly patients owing to age, educational level, and cognitive ability, among other implementation issues like the time management of physicians' firms. Novel mobile health (mHealth) platforms allow patients to communicate remotely with their health care providers. Some systems facilitate online sharing of smartphone photographs in a secure and Health Insurance Portability and Accountability Act compliant manner [7], potentially allowing for simple and rapid postoperative rehabilitation monitoring. However, building a professional mHealth platform is resource consuming and typically inefficient because patient engagement is persistently low, especially in large user groups [8].

Nevertheless, with so many people using social media every day, there is a great opportunity for mHealth initiatives to positively influence health attitudes and behaviors among large groups of people [9]. Interventions involving various media, such as radio, television, and the internet [10], have already been utilized for mass outreach health campaigns [11]. Because online social networks have several advantages, such as a large audience, increased user engagement, and high retention of contacts [12], they can strongly promote healthy behavioral changes [13]. In fact, social media approaches have already been pilot tested with promising outcomes [14,15]. Yet, studies on larger population data are pending.

WeChat (the Chinese version is *Weixin*) is the most popular instant-messaging app in China, which has been created by China's largest internet company, Tencent [16]. It is used in more than 200 countries [17], with approximately 1 billion active accounts in the first quarter of last year [18]. WeChat is a messaging and social media app where people of all ages and professions can either collaborate or just share information and messages [16]. WeChat was previously used with success in a health education program to improve malaria health literacy among Chinese expatriates and was proven to be effective, sustainable, feasible, and well accepted [17]. Similar time and cost-effectiveness were observed during the WeChat follow-up [19]. In a second example, WeChat interventions improved patient compliance and reduced the treatment duration of orthodontic treatment [2]. In this study, we scrutinized the hypothesis that social media can enhance the educational interaction between patients with joint replacement and physicians.

Methods

Study Design and Sample Selection

We performed a nonrandomized controlled study in a single-institute in-hospital population from April 2019 to January 2020. Patients receiving joint replacement surgeries in the First Affiliated Hospital of the University of Science and Technology of China were recommended to join the WeChat campaign on admission. The participants who enrolled (1) received elective lower extremity arthroplasty surgery in our joint center, (2) were willing to receive relative health and rehabilitative education, and (3) had WeChat app–compatible smartphones. Patients excluded from the study were either unwilling to participate or received emergency surgery. Participants who joined our WeChat campaign were added to a WeChat group, received at least 3 months of online follow-up, and filled a satisfaction survey. Patients not willing to join were informed about the study and were included in a control group. These patients received routine follow-up through telephone calls, phone messages, and face-to-face appointments at the clinics. They were also asked to fill a written questionnaire form on re-examination.

WeChat Campaign Construction

The group chat was created using the built-in function of the WeChat app (Figure 1A). The app allows individuals to create multiple user dialogues among their contacts. The initial group chat was built among orthopedic surgeons (n=6), ward and clinic nurses (n=3), and physical therapists (n=3). A QR code was generated for each group automatically (Figure 1B). Thereafter, the QR codes were shared with the patients in order to join the group before leaving the hospital. The maximum number of group members is 500. Therefore, a new chat group was created monthly to facilitate maintenance. The medical staff participating in the group used their real names. Patients were encouraged to use real names, although they were reassigned in-hospital IDs by the group administrator (CZ). Health advice, including clinical and hospital instructions, rehabilitation guides, and general health education, were released weekly by the doctors, nurses, and therapists. Patients raised point-to-point questions to doctors using the @ function (Figure 1C). They were also encouraged to complain or share their surgery and rehabilitation experiences (Figure 1D).

Figure 1. Creation of a chat group using the smartphone social media app WeChat. (A) Initiation of a group chat with personal contacts. (B) Generation of a QR code for patient invitation. (C) Consultation with a specific doctor/nurse/therapist using the @ function. (D). Sharing of treatment and rehabilitation experiences with video clips. Note that patient profile photos and names are hidden.

Interventions and Measures

Demographic data, such as gender, age, and educational level, were collected upon hospital admission. The patients in both groups received the same protocol treatment, rehabilitation, and education during hospital stay. After discharge, the interventions were provided through the WeChat app and routine publicities, including verbal publicity, telephone interviews, and a two-page take-home brochure. The patients were asked to fill a perioperative education and satisfaction questionnaire online or on re-examination. The questionnaire was originally developed and pilot tested by the Danish Health and Medicine Authority (Danish Health and Medicine Authority 2006) (Multimedia Appendix 1) [1]. Outcomes, including information given and overall satisfaction, were measured on a numerical rating scale from 0 (not satisfied at all) to 10 (best possible satisfaction).

WeChat Dialogue Analysis

The group chat dialogues were recorded and encrypted in the WeChat app (WeChat 7.0.10 for iOS, Tencent Co, Ltd). The records were exported in CSV format to a database using Sync tools (I4Assistant, v7.98.12, Shenzhen Waip Information Technology Co, Ltd). Information regarding group number, message timestamp, user ID (identifier), message type, and content was read from the database. The users were marked as P (patient) and D (doctor) according to the user IDs. Patients who had more than 10 single messages or three conversations were characterized as “active participants.” The message contents were recorded in plain text, figures, and videos. Voice messages were converted to text via the WeChat built-in speech recognition function. External weblinks and system messages were dropped. Text processing and analysis was preformed using Python v3.7.0 (Python Software Foundation). For English content, a blank character was used as a mandatory tokenizer. For Chinese content, we imported the external library Jieba [20] for word segmentation. The word frequency was calculated among the dialogues, and the conversation contents were clustered into different groups. A word cloud was drawn using the external library wordcloud 1.3.3 [21], and the additional Python libraries xlrd, xlwt, math, and matplotlib were used for the analysis and the corresponding plots. The drawing source code is provided in Multimedia Appendix 2.

Communication between patients was analyzed using a patient interaction score. Specifically, the first message in every group chat was assigned an interaction score value of zero, and the score of the next message was increased by one if it originated from a different patient user. However, the score was forced to zero if the message was sent from medical staff. This parameter describes how intensive the communication is among different patients under the supervision and guidance of a doctor. The group activity was determined by activeness rates defined as the total score divided by the total number of messages. The flowchart of the process is given in Multimedia Appendix 3.

Ethics Statement

The study was approved by the hospital institutional review board. All followed procedures were in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. All participants gave verbal consent to participate in the study, and they were entitled to withdraw from the study whenever they wished and for whatever reason. Private data, such as telephone numbers and identification numbers, were deleted during data analysis.

Statistical Analysis

Statistical analysis was carried out using SPSS Statistics version 22.0 (2013, SPSS Inc). For continuous variables, analysis of variance (ANOVA) and post-hoc analysis by Fisher least significant difference were applied to compare the means among the study groups. For categorical variables, logistic regression analysis and contingency tables were used. Spearman and Pearson analyses were performed to investigate monotonic and linear relationships, respectively. To control for confounding factors, propensity score methods with a multinominal logistic regression were utilized. The propensity score replaced all single covariates to adjust the effectiveness on patient satisfaction. A hierarchical regression model was used to test potential mediating or moderating effects. Data were expressed as mean (SD) for continuous variables or percentage of the total for categorical variables. A *P* value <.05 was considered statistically significant.

Results

Participant Characteristics and WeChat Activeness

A total of 3428 patients were invited to participate in the study, including 2292 in the WeChat group and 1236 in the control group. Data were collected from 2930 participants (2195 in the WeChat group and 735 in the control group). After 3 months of intervention and follow-up, information was obtained and analyzed (Figure 2). The mean age of the enrolled participants was 62.19 years (SD 12.67 years). In total, 1973 (67.34%) females and 957 (32.66%) males were included in this survey. Additionally, 2327 (79.42%) individuals had basic education, 250 (8.53%) had medial education, and 353 (12.05%) had high education. A total of 1034 (35.29%) patients underwent total hip arthroplasty, 73 (2.49%) underwent hemiarthroplasty, 96 (3.28%) underwent total hip revision, 1589 (54.23%) underwent total knee arthroplasty, 91 (3.16%) underwent unicompartmental knee arthroplasty, and 47 (1.60%) underwent total knee revision. The baseline demographics (age, gender, and educational level) were not balanced between the two groups (Table 1). Patients who had more than 10 single messages or three conversations formed the group of “active participants.” We found 738 out of 2195 (33.62%) patients active in the WeChat group (ie, the majority were inactive in the group chats). Active patients were typically males, young, and educated. The differences between the control group and WeChat group in terms of demographics were all statistically significant (Table 1).

Figure 2. Flowchart of participation.

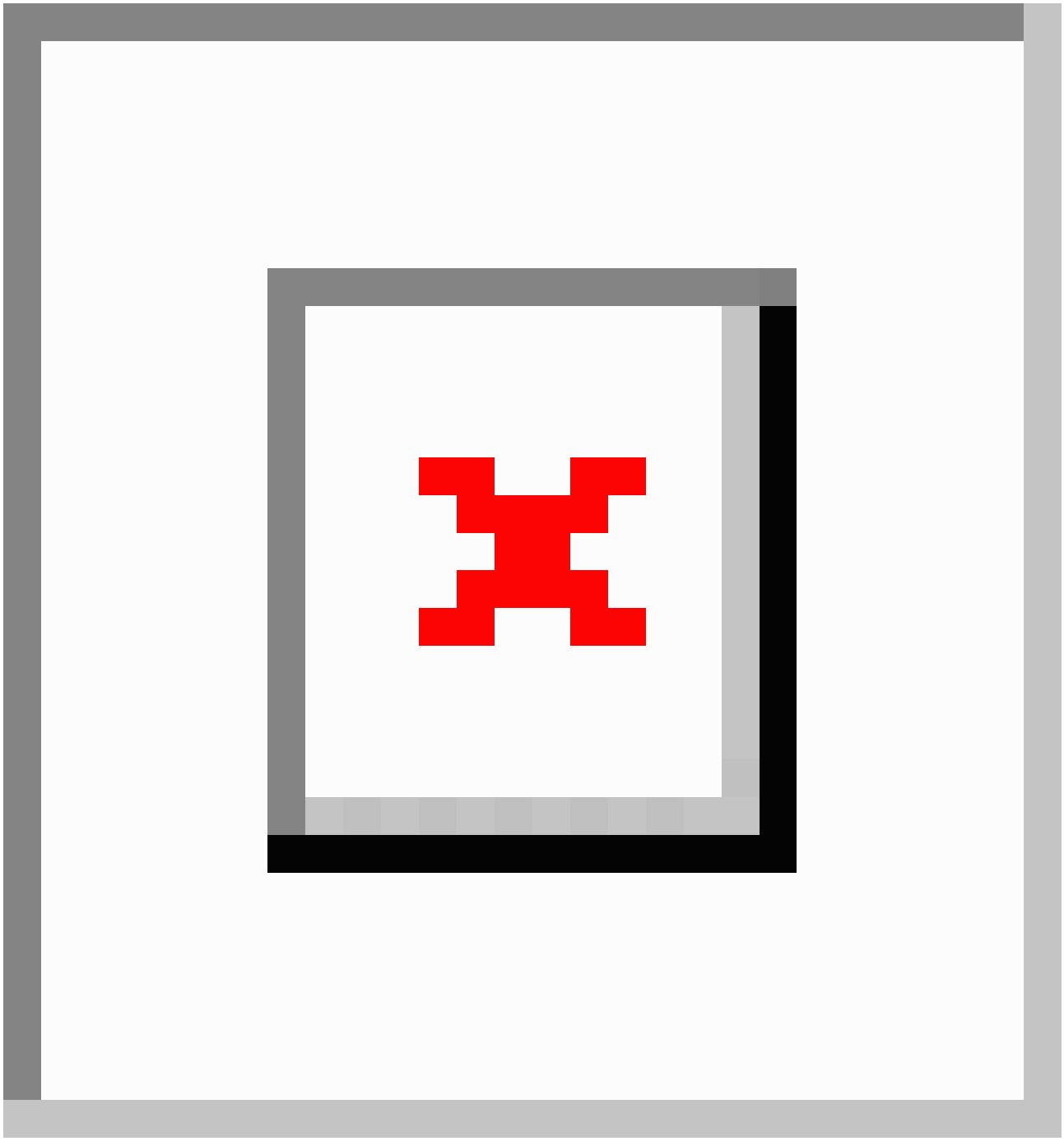


Table 1. Demographic characteristics of the control group and WeChat group (N=2930).

Demographic characteristic	Control group (n=735), n (%)	WeChat group (n=2195), n (%)			Chi-square (*df*)[a]	*P* value[a]
		Inactive (n=1457)	Active (n=738)	Total		
Gender					41.88 (2)	<.001
Male	223 (30.34)	422 (28.96)	312 (42.28)	734 (33.44)		
Female	512 (69.66)	1035 (71.04)	426 (57.72)	1461 (66.56)		
Age group (years)					400.194 (2)	<.001
<65	339 (46.12)	543 (37.27)	605 (81.98)	1148 (52.30)		
≥65	396 (53.88)	914 (62.73)	133 (18.02)	1047 (47.70)		
Education level[b]					482.017 (4)	<.001
Low	597 (81.22)	1286 (88.26)	444 (60.16)	1730 (78.82)		
Medial	87 (11.84)	123 (8.44)	40 (5.42)	163 (7.43)		
High	51 (6.94)	48 (3.29)	254 (34.42)	302 (13.76)		
Surgery type					379.55 (10)	<.001
THA[c]	162 (22.04)	410 (28.14)	462 (62.60)	872 (39.73)		
HA[d]	23 (3.13)	43 (2.95)	7 (0.95)	50 (2.28)		
THR[e]	9 (1.22)	59 (4.05)	28 (3.79)	87 (3.96)		
TKA[f]	488 (66.39)	886 (60.81)	215 (29.13)	1101 (50.16)		
UKA[g]	41 (5.58)	29 (1.99)	21 (2.85)	50 (2.28)		
TKR[h]	12 (1.63)	30 (2.06)	5 (0.67)	35 (1.60)		

[a]Chi-square and *P* values indicate differences between the control group, inactive group, and active group.

[b]Education level: low, compulsory education; medial, high school or equivalent; high, university/college or above.

[c]THA: total hip arthroplasty.

[d]HA: hemiarthroplasty.

[e]THR: total hip revision.

[f]TKA: total knee arthroplasty.

[g]UKA: unicompartmental knee arthroplasty.

[h]TKR: total knee revision.

Overall Satisfaction Scores Between the Control Group and WeChat Group

In general, patients in the WeChat group provided higher overall satisfaction scores than those in the control group (mean 8.48, SD 1.12 vs mean 6.66, SD 1.80; *P*<.001). Subgroup analysis supported this finding in both active (mean 8.47, SD 1.11 vs mean 6.66, SD 1.80; *P*<.001) and inactive patients (mean 8.49, SD 1.13 vs mean 6.66, SD 1.80; *P*<.001) (Figure 3A). Since the baseline demographics were not balanced, the correlations between overall satisfaction and gender, age, education, and surgery type were tested. It was found that overall satisfaction was not related to gender (*P*=.07), age (*P*=.21), and education level (*P*=.63), but was significantly related to surgery type (*P*<.001). Therefore, a stratified analysis was performed in order to eliminate confounding bias (Table 2). It was found that patients receiving primary lower extremity arthroplasty reported significantly higher overall satisfaction scores in the WeChat group than those in the control group. The difference between groups was not significant in patients receiving revision surgery.

Figure 3. Data analysis from the Danish Health and Medicine Authority 2006 questionnaire. (A) Histogram shows higher overall satisfaction in active patients and inactive patients in the WeChat group than in the control group. (B) Linear regression between overall satisfaction and information given. (C) The mediated effect model among information given, WeChat participation, and overall participation.

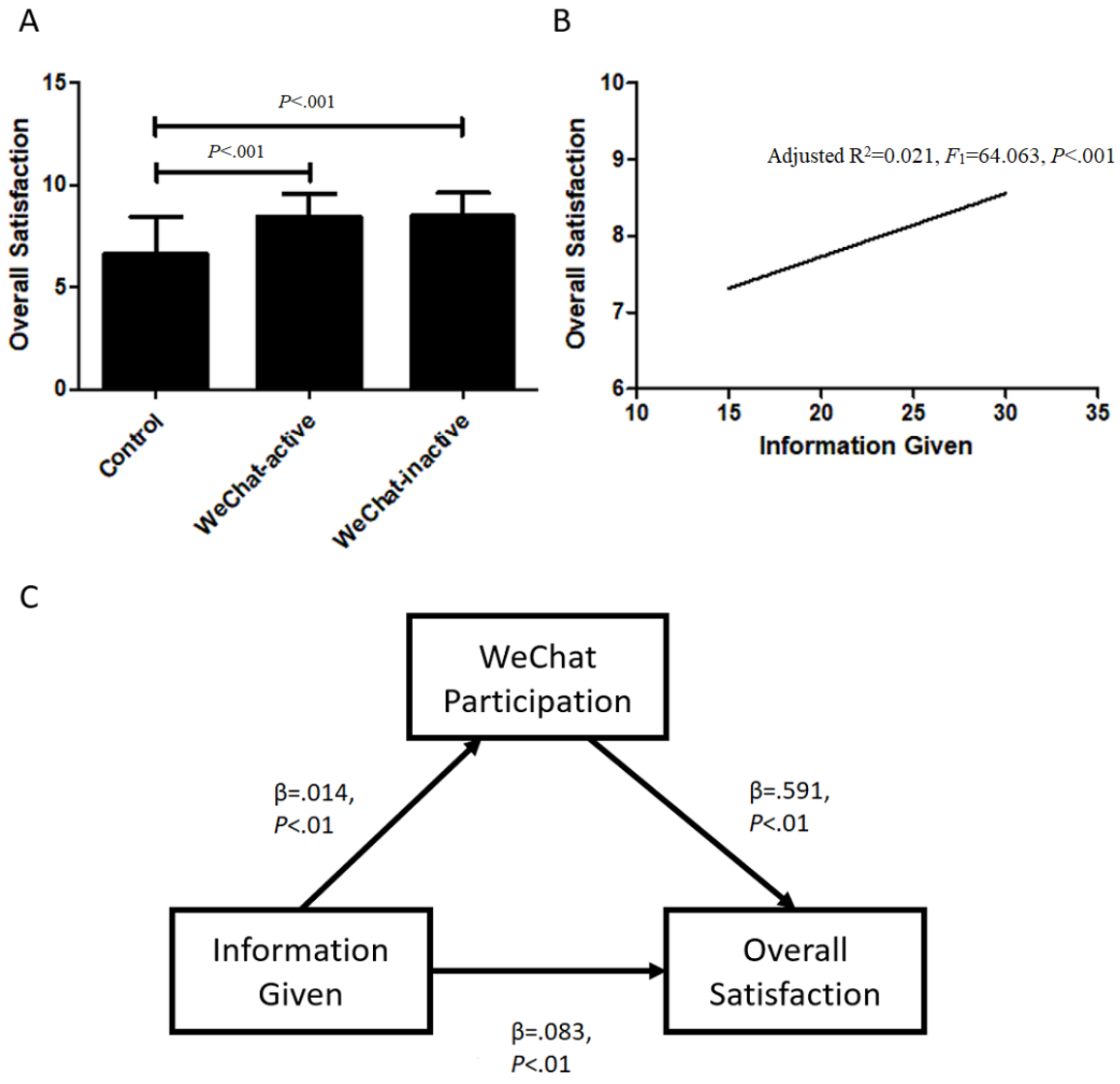


Table 2. Total satisfaction score in the control group and WeChat group by age and surgery type (N=2930).

Surgery type and group	Value, n	Score, mean (SD)	*t* (*df*)	*P* value
THA[a]			−12.525 (181.914)	<.001
Control	162	6.51 (1.94)		
WeChat	872	8.47 (1.13)		
HA[b]			−4.805 (28.914)	<.001
Control	23	6.52 (1.95)		
WeChat	50	8.62 (1.12)		
THR[c]			−1.514 (9.161)	.16
Control	9	7.78 (1.39)		
WeChat	87	8.51 (1.15)		
TKA[d]			−21.444 (667.963)	<.001
Control	488	6.65 (1.74)		
WeChat	1101	8.49 (1.11)		
UKA[e]			−4.337 (70.902)	<.001
Control	41	7.02 (1.65)		
WeChat	50	8.36 (1.19)		
TKR[f]			−2.061 (13.186)	.06
Control	12	7.17 (2.21)		
WeChat	35	8.54 (1.17)		

[a]THA: total hip arthroplasty.

[b]HA: hemiarthroplasty.

[c]THR: total hip revision.

[d]TKA: total knee arthroplasty.

[e]UKA: unicompartmental knee arthroplasty.

[f]TKR: total knee revision.

Analysis of Perioperative Education and Overall Satisfaction in Patients Receiving Arthroplasty

Perioperative education was measured using *information given* scores acquired from the Danish Health and Medicine Authority satisfaction questionnaire. A significant linear correlation was found between information given and overall satisfaction scores (adjusted R^2=0.021, F_1=64.063, P<.001) (Figure 3B). Owing to the small effect size (β=.08), we constructed and tested a causal mediation effect model to illustrate the relationship among perioperative education, WeChat campaign participation, and overall patient satisfaction. The percentage of the mediated effect of WeChat participation between perioperative education and overall patient satisfaction was 10.75% (Figure 3C).

WeChat Group Conversation Analysis

From April 2019 to March 2020, a total of 23,088 messages were recorded in eight group chats. There were 7422 (32.15%) messages from doctors and 15,666 (67.85%) messages from patients. The types of messages were as follows: text, 18,231 (78.96%); emoticons, 2226 (9.64%); figures, 2298 (9.95%); voice messages, 240 (1.04%); and videos, 93 (0.40%). The purposes of messages were as follows: medication consultation, 867 (4.69%); rehabilitation scheme, 1425 (7.72%); subsequent therapy, 3782 (20.48%); re-examination, 3769 (20.41%); reflect therapeutic effects, 3249 (17.59%); report recovery progress, 1970 (10.67%); inquire about operation related complications, 816 (4.42%); complaints, 249 (1.35%); and nonmedical information, 2344 (12.69%) (Figure 4A). The content of messages was transformed into plain text, and the resulting unstructured text of 315,900 Chinese characters was tokenized using the Python library Jieba. The top 100 words were extracted for further analysis. Nonmedical words were screened out manually in this step. The top watched words included operation, re-examination, discharge, take out stitches, and hospital. The top 30 words with the highest frequency are shown in the form of a word cloud in Figure 4B and 4C. A positive side-effect of our approach was that six hip joint dislocations and 11 surgical-site infections were detected and reported early in the WeChat patient cohort.

Apart from one-to-one conversations between doctors and patients, conversations between patients were analyzed with interaction scores and activity rates (Figure 5). The interaction intensity is shown in a heatmap in Figure 6A. Among the eight temporally continuous group chats, the highest interaction score (35 points) was found in Group 3. We also found that patients

in groups with higher activeness rates (Group 1 and Group 3) had better overall satisfaction (Figure 6B). Linear regression of overall satisfaction over the activeness rate was attempted, but it resulted in nonsignificance (adjusted R^2=0.28, F_1=3.36, P=.13).

Figure 4. Text analysis of group chat records. (A) Source, classification, and purpose of the messages. (B) Word cloud of the most frequently mentioned words in the group chats. (C) Top 30 medical words that appeared in the group chats.

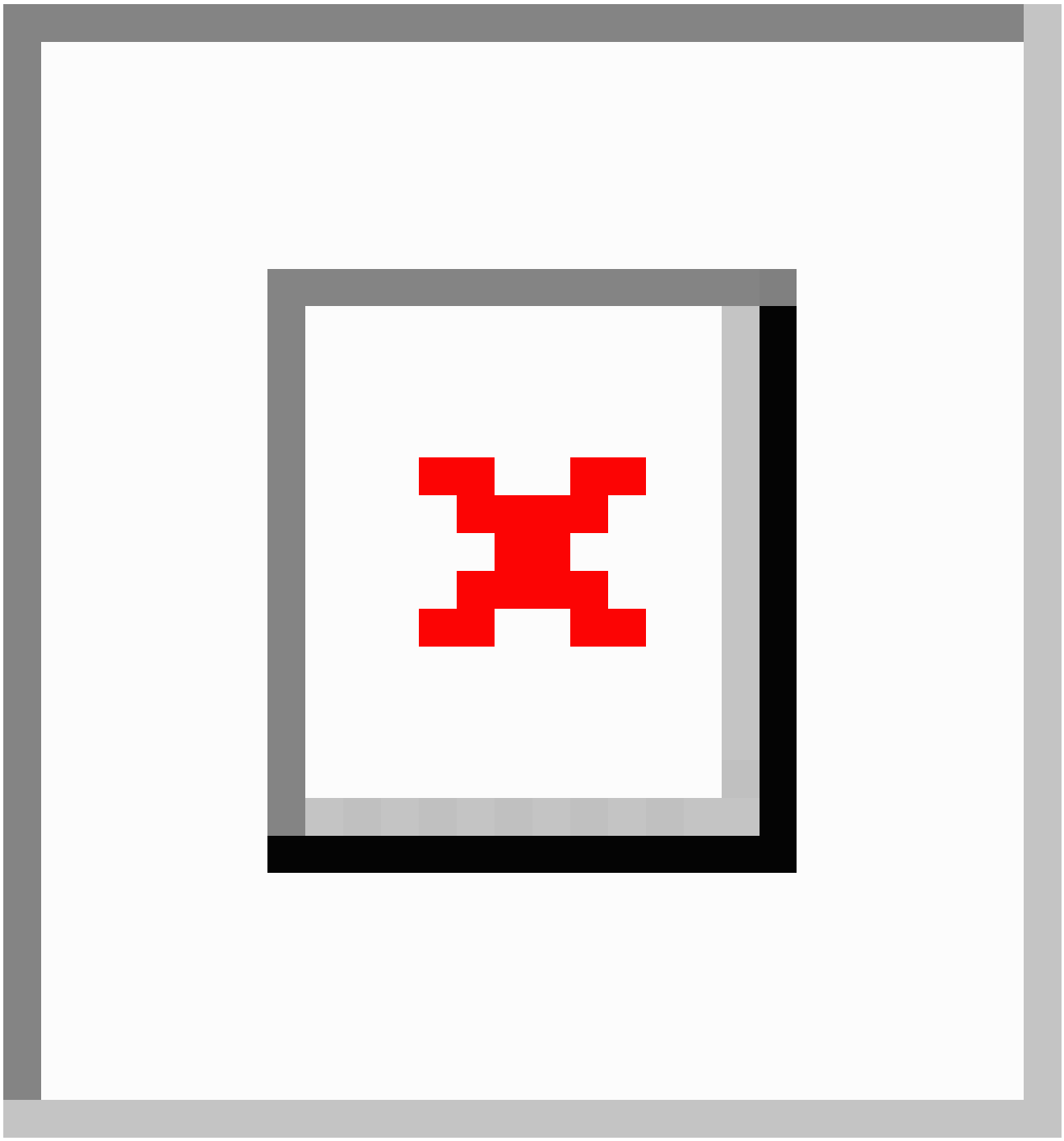


Figure 5. Doctor-patient communication and health educational model. (A) Left: doctor-patient conversation; right: patient-patient conversation in the group chat. (B) Different interaction models between doctors and patients. Left: one-to-one interactions between doctors and patients; middle: unilateral publicity from doctors to patients; right: social media–promoted multidirectional educational community among doctors and patients.

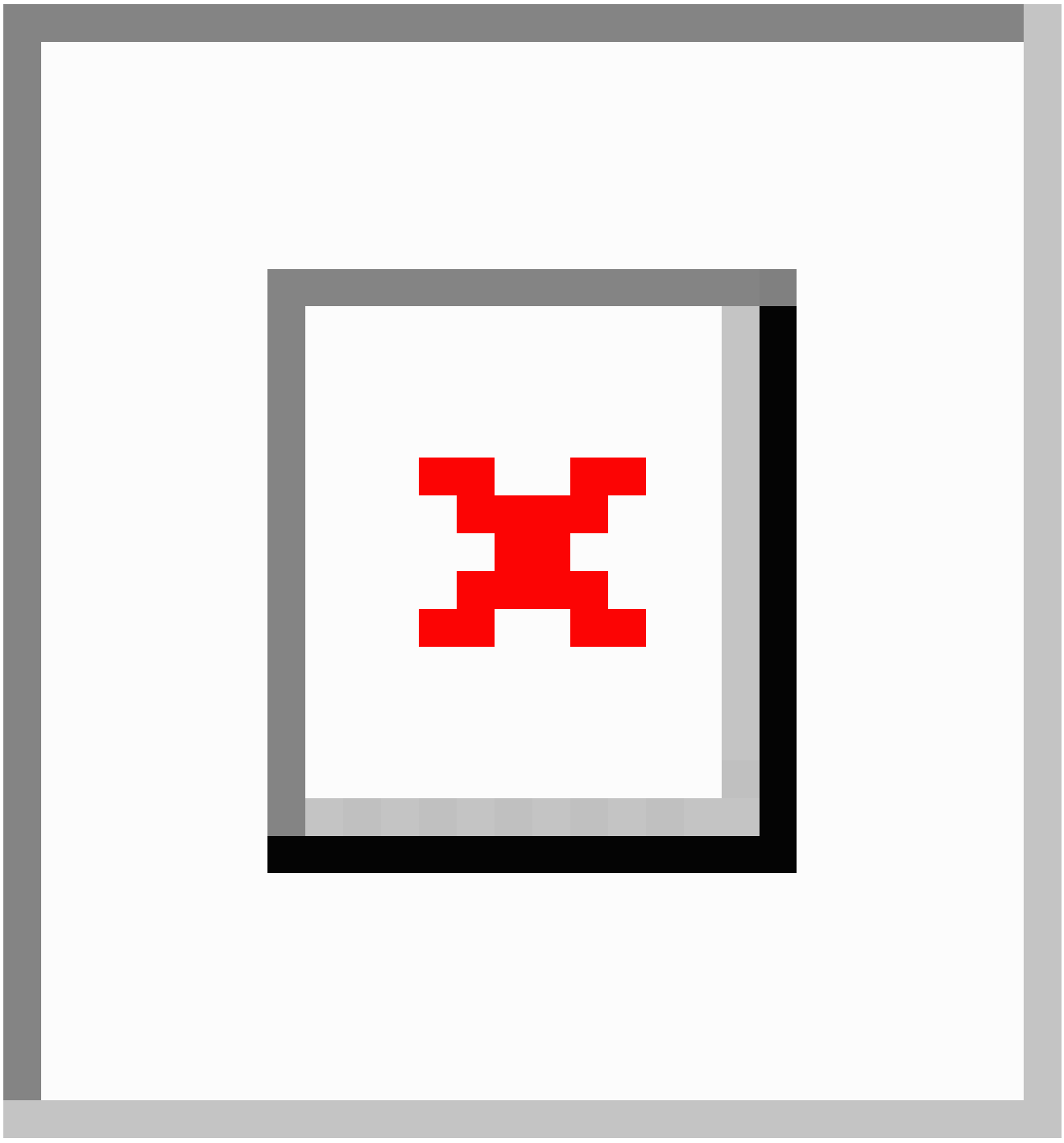
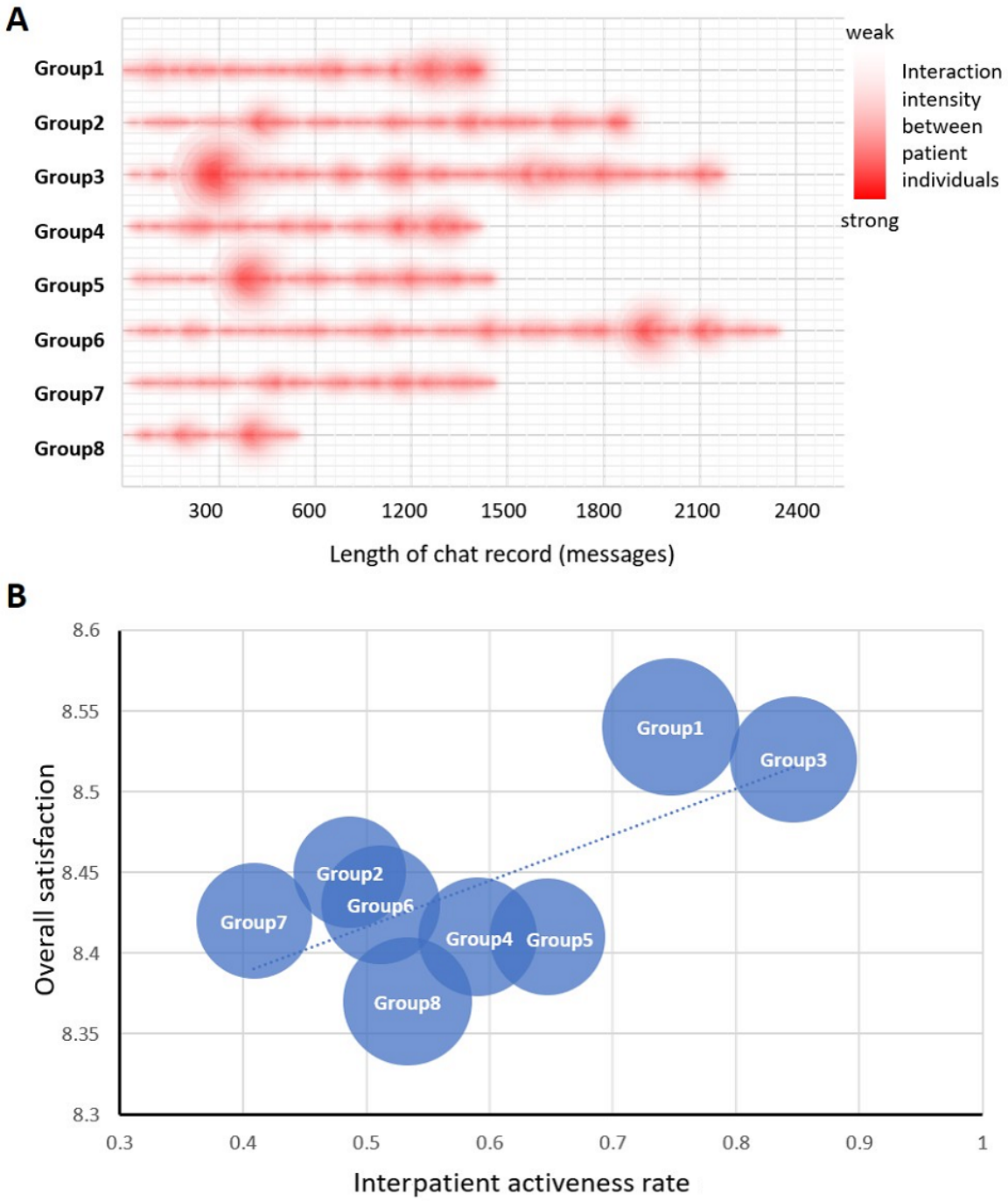


Figure 6. The patient-patient interaction intensity. (A) A heatmap reflecting the interaction intensity in different portions of the group chats. Abscissa: total length of the entire group chat. (B) Balloon map showing the overall satisfaction and interpatient activeness rate in different groups. The diameter of each circle is determined by the number of group members.



Discussion

Principal Findings

In addition to surgical success, long-term postoperative rehabilitation is essential for better outcomes after joint replacement surgery. Inadequate patient compliance may result in unsatisfactory joint function and poor quality of life in the course of home-based rehabilitation. Persistent medical intervention after discharge is important for joint function recovery and patient satisfaction [2,22]. Standard care for the rehabilitation of knee conditions involves exercise programs and information provision. Current methods of rehabilitation struggle to keep up with the large volumes of patients and the lengthy treatment required for recovery. Herein, we described a novel smartphone-based community approach for rehabilitation guidance and medical support in postoperative patients. This nonrandomized study showed that a social media–enhanced educational community is an effective

approach to promote patients' satisfaction with primary arthroplasty.

Social media can influence lifestyle, including health behavior [23], especially for large groups of people [24]. In our study, 2292 out of 3428 (66.86%) patients who underwent joint replacement surgery were willing to participate in the WeChat-based community to evaluate the quality and the effect of postoperative care via the internet. Compared with professional mHealth platforms, the social media WeChat-based community has far more active users. It is estimated that online social networks account for approximately 27.18% of all the time people spend online [3], and 55.2% of WeChat users check the WeChat status over 10 times per day, while 25.0% of users check the status over 30 times per day [25]. Therefore, we hypothesized that medical information can be delivered more efficiently in this way.

In general, participants in the WeChat group had higher overall satisfaction scores (mean 8.48, SD 1.12) than those in the control group (mean 6.66, SD 1.80), which is consistent with previous studies [3,4,19]. The smartphone WeChat app has been found to be a viable option for follow-up in discharged patients with head and neck tumors [19]. Therefore, the doctor-led follow-up model has the potential to establish a good physician-patient relationship by enhancing dynamic communication and providing individual health instructions. Therapeutic guidance and intervention via WeChat also improved rehabilitation after total hip arthroplasty and promoted the recovery of joint function in patients [3]. The platform helps patients to comprehend disease knowledge and rehabilitation exercise methods, promoting recovery [4]. Using stratified analysis, we observed the significant satisfaction advantages of WeChat in patients receiving primary lower limb extremity arthroplasty, while the effect was less notable in patients receiving revision surgeries. Despite the fact that the long-term quality of life is poorer after revision surgery than after primary surgery [26], the patient satisfaction scores were relatively high in the WeChat and control groups during hospitalization, probably because patients who had a previous operation history were better informed and less anxious.

Perioperative patient education is widely used to inform patients and relatives about various aspects of the upcoming operation and to motivate patients to be active participants [6]. However, the effects of perioperative patient education on surgical outcomes remain to be considered. Louw et al found that adding a brief 30-minute pain neuroscience education session to a traditional preoperative total knee arthroplasty education program did not result in any significant improvements, except for patient satisfaction [6]. Perioperative patient education might increase patient satisfaction by communicating realistic expectations to the patients [5]. In this study, a linear regression between overall patient satisfaction and information given was observed as an indicator for perioperative patient education levels. Although statistically significant (P<.001), the effect size of perioperative patient education on overall satisfaction was small (β=.08). It is most likely that WeChat participation mediates a perioperative patient education effect on patient satisfaction (ie, the more information given, the more participation is expected in the community interaction).

An important question is "how does WeChat participation and activeness influence patient satisfaction?" This study showed that although patients in the WeChat group had higher overall satisfaction scores than those in the control group, the scores between active and inactive patients were indifferent (Figure 3A). It was doubted whether this difference came from the selection bias due to the nonrandomized study design. However, on testing the baseline demographics, the difference remained insignificant between the control and WeChat groups. Instead, young and highly educated participants were among WeChat active users in contrast to inactive users. We believe that inactive users, although not participating in the group chat, still received information and social support from the experience shared by other active users. A series of reports has proven the positive effect of social support on the outcomes of lower limb extremity arthroplasty [27-30]. In studies of patients with hip fractures, a significant relationship was also identified between social support and lower limb functional activity [31]. Sveikata et al demonstrated better postoperative functional results 12 months after total knee arthroplasty in patients who had better social support [32].

Added social support stems from the interpatient interaction in the group chats. In addition to the traditional doctor-led follow-up models, medical doctors typically communicate with a single patient or address the public unilaterally. However, this new social media approach enables patients to play an active role in the network (Figure 6). Our study showed that patients in groups with more interpatient interaction were more satisfied with their overall treatment. The structural support may provide individuals with self-respect and motivation [28] for more appropriate health decisions [33]. Having a large amount of structural support from group chat members may also mean greater access to sources of information, which can increase the likelihood of having access to accurate and relevant information sources [28]. Functional support, on the other hand, may have a positive effect on health decision-making during periods of distress [30]. When an individual perceives high levels of social support, he or she may reassess a stressful event as less worrisome and become more capable to face a disease [11]. Therefore, it is important to know about the rehabilitation experience and to receive suggestions from other patients during recovery.

The effectiveness of medical interventions involving smartphone apps compared with those involving the internet has been debated, and the latter approaches have been suggested as being more accessible and convenient for patients [34], although a more recent report found smartphone apps to be more efficient and effective for online health care services [35], possibly due to the tremendous growth of smartphone usage. In the context of physical activity, it was found that over half of the controlled trials of web-based interventions reported positive behavioral outcomes [36]. Furthermore, a web-based intervention was found to be useful in delivering standard care for the rehabilitation of knee conditions [15]. However, poor adherence is a common problem with web-based approaches [15,37]. On the other hand, the most effective interventions are interactive and flexible, thereby allowing patients to select information that is relevant to their medical conditions and suitable at a particular

point in time [37]. Facebook has been already used as a tool to promote public health and educational health services [38,39], while Twitter has been, to some extent, a possible alternative [40].

To date, social media usage is still controversial owing to privacy, security, confidentiality, and liability issues [15]. According to laws and regulations, communication between health care providers and patients must comply with current data security and privacy legislation. The presented social media community approach has the legal risk of leaking private patient information in an open group chat. Health care–centric discussions must ensure trusted moderation and storage facilities. We have to also pay attention to avoiding the possibility of spreading fake news and wrong interventions [41]. The medical staff members in the group therefore have the responsibility of community supervision.

This study has some limitations. First, it was a single-institute nonrandomized controlled study. Although propensity score methods were used and stratified analysis was performed, the conclusions were still limited to a small data set. Second, it was a retrospective study, and the baseline data (eg, education level) of the patients were not consistent between the different groups. Some patients may refuse to join the social media campaign simply because of electronic illiteracy. Additionally, a difference in satisfaction would be unsurprising. Further prospective balance-armed studies should be conducted to eliminate this issue. Third, patient satisfaction is a complex variable affected by multiple noncontrollable factors that were not evaluated (probably cannot be evaluated). Although this study is very preliminary, it is one of the few studies to analyze the interactions among patients for social media education after lower extremity joint replacement. Further investigations and clinical trials will bring more value to this field.

Conclusions

The present social media–promoted educational community based on WeChat can improve overall satisfaction among hip and knee joint replacement patients. This new model of physician-patient community interaction can facilitate better postoperative rehabilitation.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant number: 81871788 and 81902201), the Project for Science and Technology Leader of Anhui Province (grant number: 2018H177), the Scientific Research Fund of Anhui Education (grant number: 2017jyxm1097), and the Anhui Provincial Postdoctoral Science Foundation (grant number: 2019B302).

Authors' Contributions

XZ and CZ conceived the idea and designed the project. XC performed the data analysis. RG, XZ, and CZ wrote the paper. XZ, GL, and NK revised the manuscript. CZ and GL obtained funding support. All authors have read and approved the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The questions used in the Danish Health and Medicine Authority 2006 questionnaire.
[DOCX File , 16 KB - mhealth_v9i3e18763_app1.docx]

Multimedia Appendix 2
The drawing source code to build the word cloud image.
[TXT File , 0 KB - mhealth_v9i3e18763_app2.txt]

Multimedia Appendix 3
The pseudocode flowchart to calculate the interaction score for messages.
[PNG File , 79 KB - mhealth_v9i3e18763_app3.png]

References

1. Husted H, Hansen HC, Holm G, Bach-Dal C, Rud K, Andersen KL, et al. [Accelerated versus conventional hospital stay in total hip and knee arthroplasty III: patient satisfaction]. Ugeskr Laeger 2006 May 29;168(22):2148-2151. [Medline: 16768952]
2. Li X, Xu Z, Tang N, Ye C, Zhu X, Zhou T, et al. Effect of intervention using a messaging app on compliance and duration of treatment in orthodontic patients. Clin Oral Investig 2016 Nov 2;20(8):1849-1859. [doi: 10.1007/s00784-015-1662-6] [Medline: 26631059]
3. Luo J, Dong X, Hu J. Effect of nursing intervention via a chatting tool on the rehabilitation of patients after Total hip Arthroplasty. J Orthop Surg Res 2019 Dec 09;14(1):417 [FREE Full text] [doi: 10.1186/s13018-019-1483-4] [Medline: 31818330]

4. Wang J, Tong Y, Jiang Y, Zhu H, Gao H, Wei R, et al. The effectiveness of extended care based on Internet and home care platform for orthopaedics after hip replacement surgery in China. J Clin Nurs 2018 Nov 01;27(21-22):4077-4088. [doi: 10.1111/jocn.14545] [Medline: 29851157]
5. Ghomrawi HMK, Lee L, Nwachukwu BU, Jain D, Wright T, Padgett D, et al. Preoperative Expectations Associated With Postoperative Dissatisfaction After Total Knee Arthroplasty. Journal of the American Academy of Orthopaedic Surgeons 2020;28(4):e145-e150. [doi: 10.5435/jaaos-d-18-00785]
6. Louw A, Puentedura EJ, Reed J, Zimney K, Grimm D, Landers MR. A controlled clinical trial of preoperative pain neuroscience education for patients about to undergo total knee arthroplasty. Clin Rehabil 2019 Nov 19;33(11):1722-1731. [doi: 10.1177/0269215519857782] [Medline: 31213078]
7. Zhang J, Dushaj K, Rasquinha VJ, Scuderi GR, Hepinstall MS. Monitoring Surgical Incision Sites in Orthopedic Patients Using an Online Physician-Patient Messaging Platform. J Arthroplasty 2019 Sep;34(9):1897-1900. [doi: 10.1016/j.arth.2019.05.003] [Medline: 31186183]
8. He C, Wu S, Zhao Y, Li Z, Zhang Y, Le J, et al. Social Media-Promoted Weight Loss Among an Occupational Population: Cohort Study Using a WeChat Mobile Phone App-Based Campaign. J Med Internet Res 2017 Oct 23;19(10):e357 [FREE Full text] [doi: 10.2196/jmir.7861] [Medline: 29061555]
9. Noar SM. A 10-year retrospective of research in health mass media campaigns: where do we go from here? J Health Commun 2006 Jan;11(1):21-42. [doi: 10.1080/10810730500461059] [Medline: 16546917]
10. Davies CA, Spence JC, Vandelanotte C, Caperchione CM, Mummery W. Meta-analysis of internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012 Apr 30;9(1):52 [FREE Full text] [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]
11. Hunter RF, de la Haye K, Murray JM, Badham J, Valente TW, Clarke M, et al. Social network interventions for health behaviours and outcomes: A systematic review and meta-analysis. PLoS Med 2019 Sep 3;16(9):e1002890 [FREE Full text] [doi: 10.1371/journal.pmed.1002890] [Medline: 31479454]
12. De Bruyn A, Lilien GL. A multi-stage model of word-of-mouth influence through viral marketing. International Journal of Research in Marketing 2008 Sep;25(3):151-163. [doi: 10.1016/j.ijresmar.2008.03.004]
13. Maher CA, Lewis LK, Ferrar K, Marshall S, De Bourdeaudhuij I, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014 Feb 14;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
14. Tatla SK, Shirzad N, Lohse KR, Virji-Babul N, Hoens AM, Holsti L, et al. Therapists' perceptions of social media and video game technologies in upper limb rehabilitation. JMIR Serious Games 2015 Mar 10;3(1):e2 [FREE Full text] [doi: 10.2196/games.3401] [Medline: 25759148]
15. Spasić I, Button K, Divoli A, Gupta S, Pataky T, Pizzocaro D, et al. TRAK App Suite: A Web-Based Intervention for Delivering Standard Care for the Rehabilitation of Knee Conditions. JMIR Res Protoc 2015 Oct 16;4(4):e122 [FREE Full text] [doi: 10.2196/resprot.4091] [Medline: 26474643]
16. Zeng F, Deng G, Wang Z, Liu L. WeChat: a new clinical teaching tool for problem-based learning. Int J Med Educ 2016 Apr 25;7:119-121 [FREE Full text] [doi: 10.5116/ijme.5708.e5c4] [Medline: 27111920]
17. Li W, Han LQ, Guo YJ, Sun J. Using WeChat official accounts to improve malaria health literacy among Chinese expatriates in Niger: an intervention study. Malar J 2016 Nov 24;15(1):567 [FREE Full text] [doi: 10.1186/s12936-016-1621-y] [Medline: 27881122]
18. Sun M, Yang L, Chen W, Luo H, Zheng K, Zhang Y, et al. Current status of official WeChat accounts for public health education. J Public Health (Oxf) 2020 Jan 23. [doi: 10.1093/pubmed/fdz163] [Medline: 31974552]
19. Lyu KX, Zhao J, Wang B, Xiong GX, Yang WQ, Liu QH, et al. Smartphone Application WeChat for Clinical Follow-up of Discharged Patients with Head and Neck Tumors: A Randomized Controlled Trial. Chin Med J (Engl) 2016 Dec 05;129(23):2816-2823 [FREE Full text] [doi: 10.4103/0366-6999.194635] [Medline: 27900995]
20. jieba. GitHub. URL: https://github.com/fxsjy/jieba [accessed 2021-02-21]
21. wordcloud. Python Software Foundation. URL: https://pypi.org/project/wordcloud/ [accessed 2021-02-21]
22. Shortell SM, Rundall TG, Hsu J. Improving patient care by linking evidence-based medicine and evidence-based management. JAMA 2007 Aug 08;298(6):673-676. [doi: 10.1001/jama.298.6.673] [Medline: 17684190]
23. Guo Y, Hong YA, Cai W, Li L, Hao Y, Qiao J, et al. Effect of a WeChat-Based Intervention (Run4Love) on Depressive Symptoms Among People Living With HIV in China: A Randomized Controlled Trial. J Med Internet Res 2020 Feb 09;22(2):e16715 [FREE Full text] [doi: 10.2196/16715] [Medline: 32044751]
24. Zhang Y, Xia T, Huang L, Yin M, Sun M, Huang J, et al. Factors Influencing User Engagement of Health Information Disseminated by Chinese Provincial Centers for Disease Control and Prevention on WeChat: Observational Study. JMIR Mhealth Uhealth 2019 Jun 27;7(6):e12245 [FREE Full text] [doi: 10.2196/12245] [Medline: 31250833]
25. WeChat User Global Report 2015. TMTPost. URL: https://www.tmtpost.com/1021929.html [accessed 2021-03-03]
26. Patil S, Garbuz DS, Greidanus NV, Masri BA, Duncan CP. Quality of life outcomes in revision vs primary total hip arthroplasty: a prospective cohort study. J Arthroplasty 2008 Jun;23(4):550-553. [doi: 10.1016/j.arth.2007.04.035] [Medline: 18514873]

27. Kiefer RA. The effect of social support on functional recovery and wellbeing in older adults following joint arthroplasty. Rehabil Nurs 2011;36(3):120-126. [doi: 10.1002/j.2048-7940.2011.tb00077.x] [Medline: 21675398]
28. Vina ER, Cloonan YK, Ibrahim SA, Hannon MJ, Boudreau RM, Kwoh CK. Race, sex, and total knee replacement consideration: role of social support. Arthritis Care Res (Hoboken) 2013 Jul 01;65(7):1103-1111 [FREE Full text] [doi: 10.1002/acr.21925] [Medline: 23281259]
29. Wu K, Lee P, Chou W, Chen S, Huang Y. Relationship between the social support and self-efficacy for function ability in patients undergoing primary hip replacement. J Orthop Surg Res 2018 Jun 18;13(1):150 [FREE Full text] [doi: 10.1186/s13018-018-0857-3] [Medline: 29914520]
30. Zeppieri KE, Butera KA, Iams D, Parvataneni HK, George SZ. The Role of Social Support and Psychological Distress in Predicting Discharge: A Pilot Study for Hip and Knee Arthroplasty Patients. J Arthroplasty 2019 Nov;34(11):2555-2560. [doi: 10.1016/j.arth.2019.06.033] [Medline: 31327646]
31. Mock C, MacKenzie E, Jurkovich G, Burgess A, Cushing B, deLateur B, et al. Determinants of disability after lower extremity fracture. J Trauma 2000 Dec;49(6):1002-1011. [doi: 10.1097/00005373-200012000-00005] [Medline: 11130480]
32. Sveikata T, Porvaneckas N, Kanopa P, Molyte A, Klimas D, Uvarovas V, et al. Age, Sex, Body Mass Index, Education, and Social Support Influence Functional Results After Total Knee Arthroplasty. Geriatr Orthop Surg Rehabil 2017 Jun 22;8(2):71-77 [FREE Full text] [doi: 10.1177/2151458516687809] [Medline: 28540111]
33. Midboe AM. Worry and Health Decision-Making: The Moderating Role of Social Support. 2010. URL: http://d-scholarship.pitt.edu/6484/1/MidboeAmandaM_etd2010_rev.pdf [accessed 2020-01-01]
34. Luthans F, Avey JB, Patera JL. Experimental Analysis of a Web-Based Training Intervention to Develop Positive Psychological Capital. AMLE 2008 Jun;7(2):209-221. [doi: 10.5465/amle.2008.32712618]
35. Schueller SM, Washburn JJ, Price M. Exploring Mental Health Providers' Interest in Using Web and Mobile-Based Tools in their Practices. Internet Interv 2016 May;4(2):145-151 [FREE Full text] [doi: 10.1016/j.invent.2016.06.004] [Medline: 28090438]
36. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions a review of the literature. Am J Prev Med 2007 Jul;33(1):54-64. [doi: 10.1016/j.amepre.2007.02.041] [Medline: 17572313]
37. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
38. Kite J, Foley BC, Grunseit AC, Freeman B. Please Like Me: Facebook and Public Health Communication. PLoS One 2016 Sep 15;11(9):e0162765 [FREE Full text] [doi: 10.1371/journal.pone.0162765] [Medline: 27632172]
39. Zaidan AA, Zaidan BB, Kadhem Z, Larbani M, Lakulu MB, Hashim M. Challenges, alternatives, and paths to sustainability: better public health promotion using social networking pages as key tools. J Med Syst 2015 Feb 29;39(2):7. [doi: 10.1007/s10916-015-0201-y] [Medline: 25631841]
40. Gabarron E, Wynn R. Use of social media for sexual health promotion: a scoping review. Glob Health Action 2016 Sep 19;9(1):32193 [FREE Full text] [doi: 10.3402/gha.v9.32193] [Medline: 27649758]
41. Bovet A, Makse HA. Influence of fake news in Twitter during the 2016 US presidential election. Nat Commun 2019 Jan 02;10(1):7 [FREE Full text] [doi: 10.1038/s41467-018-07761-2] [Medline: 30602729]

Edited by G Eysenbach; submitted 17.03.20; peer-reviewed by Z Shi, H Liang, K Chen, D Chen, A Kulakli; comments to author 12.06.20; revised version received 25.09.20; accepted 02.10.20; published 18.03.21.

Please cite as:
Zhang X, Chen X, Kourkoumelis N, Gao R, Li G, Zhu C
A Social Media–Promoted Educational Community of Joint Replacement Patients Using the WeChat App: Survey Study
JMIR Mhealth Uhealth 2021;9(3):e18763
URL: https://mhealth.jmir.org/2021/3/e18763
doi:10.2196/18763
PMID:33734094

Tutorial

Challenges of and Solutions for Developing Tailored Video Interventions That Integrate Multiple Digital Assets to Promote Engagement and Improve Health Outcomes: Tutorial

Camilla Harshbarger[1], PhD; Olivia Burrus[2], MPH; Sivakumar Rangarajan[3], MS; John Bollenbacher[4], MA; Brittany Zulkiewicz[2], BA; Rohit Verma[5], MSc; Carla A Galindo[1], MPH; Megan A Lewis[6], PhD

[1]Division of HIV/AIDS Prevention, Centers for Disease Control and Prevention, Atlanta, GA, United States

[2]Center for Communication Science, Research Triangle International, Research Triangle Park, NC, United States

[3]Public Health Division, SeKON Enterprise Inc, Atlanta, GA, United States

[4]Critical Focus Creative, Durham, NC, United States

[5]Office of the Associate Director for Communication, Centers for Disease Control and Prevention, Atlanta, GA, United States

[6]Translational Health Sciences Division, Research Triangle International, Seattle, WA, United States

Corresponding Author:
Camilla Harshbarger, PhD
Division of HIV/AIDS Prevention, Centers for Disease Control and Prevention
1600 Clifton Road
MS8-5
Atlanta, GA
United States
Phone: 1 404 639 4267
Email: uzz9@cdc.gov

Abstract

Background: Video is a versatile and popular medium for digital health interventions. As mobile device and app technology advances, it is likely that video-based interventions will become increasingly common. Although clinic waiting rooms are complex and busy environments, they offer the opportunity to facilitate engagement with video-based digital interventions as patients wait to see their providers. However, to increase efficiency in public health, leverage the scalability and low cost of implementing digital interventions, and keep up with rapidly advancing technology and user needs, more design and development guidance is needed for video-based tailored interventions.

Objective: We provide a tutorial for digital intervention researchers and developers to efficiently design and develop video-based tailored digital health interventions. We describe the challenges and solutions encountered with Positive Health Check (PHC), a hybrid app used to deliver a brief, interactive, individually tailored video-based HIV behavioral counseling intervention. PHC uses video clips and multimedia digital assets to deliver intervention content, including interactive tailored messages and graphics, a repurposed animated video, and patient and provider handouts generated in real time by PHC.

Methods: We chronicle multiple challenges and solutions for the following: (1) using video as a medium to enhance user engagement, (2) navigating the complexity of linking a database of video clips with other digital assets, and (3) identifying the main steps involved in building an app that will seamlessly deliver to users individually tailored messages, graphics, and handouts.

Results: We leveraged video to enhance user engagement by featuring “video doctors,” full-screen video, storyboards, and streamlined scripts. We developed an approach to link the database of video clips with other digital assets through script coding and flow diagrams of algorithms to deliver a tailored user experience. We identified the steps to app development by using keyframes to design the integration of video and digital assets, using agile development methods to gather iterative feedback from multidisciplinary teams, and creating an intelligent data-driven back-end solution to tailor message delivery to individual users.

Conclusions: Video-based digital health interventions will continue to play an important role in the future of HIV prevention and treatment, as well as other clinical health practices. However, facilitating the adoption of an HIV video intervention in HIV clinical settings is a work in progress. Our experience in designing and developing PHC presented unique challenges due to the extensive use of a large database of videos tailored individually to each user. Although PHC focuses on promoting the health and

well-being of persons with HIV, the challenges and solutions presented in this tutorial are transferable to the design and development of video-based digital health interventions focused on other areas of health.

(JMIR Mhealth Uhealth 2021;9(3):e21128) doi:10.2196/21128

KEYWORDS

HIV video intervention; patient-provider communication; ART adherence; digital interventions; mobile interventions; computer based interventions; interactive technologies

Introduction

In this article, we present a tutorial on the challenges and solutions for designing and developing Positive Health Check (PHC), a brief and interactive digital video counseling intervention that can be used on computers or mobile devices (tablets, laptops) for persons with HIV who are engaged in clinical care. To improve health outcomes, the US public health community is leveraging the nation's ubiquitous use of digital media [1]. Health interventions delivered through tablets, smartphones, and computers—including telehealth and wearable and remote monitoring devices—have been shown to improve health outcomes in multiple areas, including cancer prevention and control [2], alcohol intake and physical activity [3], sedentary behavior [4], and diabetes [5]. Given the ease of broad dissemination, digital interventions offer scalability and therefore the potential for greater population health impact. Scalability is a recent focus of HIV prevention efforts by the Centers for Disease Control and Prevention (CDC) through high-impact prevention [6], a strategy calling for effective, low-cost, and scalable HIV interventions, including those that can be disseminated via the internet [7].

The well-documented surge in the development of digital interventions for HIV prevention and care [8,9] aims to prevent the transmission of HIV and enhance the health of people with HIV through secondary prevention and management [10]. HIV digital interventions are appealing because of their ability to reach specific populations and social networks at high risk for HIV transmission [11-14], often through SMS text messaging [15] and social media [16]. Additionally, educational videos have been proven to improve knowledge, attitudes, and long-term behavior change in diverse populations [17], as seen in the precedent for implementing video interventions in sexually transmitted disease (STD) clinic environments [17-20] and other health care settings [2]. Through the implementation of video-based digital interventions in the waiting room or examination room, the clinic waiting time for patients with HIV can be used as a "teachable moment" [21].

Consequently, guidance is needed to help researchers and developers design, develop, and disseminate video-based digital interventions for HIV and other fields [2,22]. This tutorial addresses technical questions around the multiple challenges and solutions for the following: (1) using video as a medium to enhance user engagement with behavioral counseling interventions, (2) navigating the complexity of linking video with other digital assets through script coding and flow diagrams of tailoring algorithms, and (3) identifying main steps to building an app that will seamlessly deliver individually tailored messages, graphics, and handouts to users.

Overview of PHC

Theory and Rationale Underlying PHC

PHC's primary aim is to enhance HIV viral suppression by providing individually tailored video messages on relevant topics, which are delivered by virtual doctors (played by actors) and mimic a clinical encounter. PHC's secondary aims are to engage people with HIV in clinical care more regularly and to reduce HIV transmission risk. The theoretical foundations of PHC include the Information-Behavior-Motivation model [23], motivational interviewing [24], and the Transtheoretical model [25]. The PHC intervention design is based on similar video-based interventions that demonstrated efficacy for self-reported antiretroviral therapy (ART) adherence [26,27], reducing sex risk [28], and reducing viral load [27]. However, the scope of PHC goes beyond previous HIV video-based interventions in that it covers the above topics and also includes modules on treatment initiation, retention in medical care, safe injection drug use, and preventing mother-to-child transmission. The formative work that guided the development of PHC is described elsewhere [29].

PHC is also designed to strengthen patient-provider communication and empower patients to voice their concerns and ask questions in medical appointments. This supports current trends in patient education, where the patient and provider collaborate to make treatment decisions [30]. The video modules (such as adherence to medication) deliver tailored motivational messages, then users select questions to ask their provider and behavior change strategies. Behavior change strategies include, for example, creating reminders to take medications or getting tested for hepatitis C. These personal choices can make the intervention more motivating and increase engagement [31]. At the end of the intervention, the user receives a personalized handout documenting their selected questions and behavior change strategies. The handouts aim to strengthen the patient-provider relationship by encouraging providers to respond to the patient's questions, which aligns with research that shows that patients may have better health outcomes when they feel their provider cares about them [32]. The behavior change strategies are for the patient to practice before the next clinic visit. In addition, the behavior change strategies are to be used as talking points to discuss with the provider, which gives the provider the opportunity to encourage positive behavior change, provide accountability, and give emotional support. Research shows that patients are more likely to sustain long-term behavior change if the patient makes an active commitment to do so [33]. Additionally, follow-up by the patient's provider may signal trust in the intervention [31]. The rationale for this intervention structure is to provide a level of personal

engagement with the tool that could lead patients from one stage of change to another (eg, from contemplating behavior change to taking action to change behavior).

Rationale for a Video and Digital Platform

Multiple design considerations aimed to maximize PHC's digital platform. First, we chose video for its visually engaging qualities and the ability to create a personal experience by tailoring the delivery of video clips to each individual user. The PHC video modules share information through visual, verbal, and written cues (on graphic overlays) and closed captioning. Second, the digital platform facilitates easy updates to PHC video clips and content in the supplemental resources. Third, video clips can be repurposed. For example, the animated condom video [34] featured in PHC that teaches how to use condoms correctly was repurposed from the video Safe in the City [20], an STD prevention intervention. Similarly, with permission, PHC video clips and the optional Extra Info microsite content featured at the end of the tool can be embedded in other interventions or used for discussion in face-to-face interventions (Multimedia Appendices 1-3). Fourth, the digital platform facilitated the use of a variety of digital assets, including tailored messages delivered via interactive video or graphics, the animated Safe in the City video, closed captioning, and user-generated patient/provider handouts. Fifth, we chose video in anticipation of meeting the demands of future users [2], as forecasts project that by 2022—along with accelerated use of Wi-Fi hotspots and increasing broadband speed—video will occupy about 84% of internet traffic [35].

How PHC Architecture Works

PHC was designed to be used in the clinic setting, where staff provide earbuds and assign a tablet with a privacy screen to the patient. To use the intervention, patients log on to a device, answer four demographic tailoring questions, and then answer risk behavior assessment questions at the start of each intervention module. Based on this information, PHC generates tailored messages. A clip of the intervention that depicts an excerpt from the adherence module is available online [36].

PHC is interactive and allows the patient to choose their "video doctor" by race and gender; if desired, they can switch doctors at every use. In each session, the patient can select intervention content and drill down into the topics they wish to learn about. For example, under the broad topic of adherence, the patient can choose to learn about side effects of medication or taking ART and drinking alcohol. PHC delivers tailored messages to each patient, drawing from a database of 700 unique video clips. Additionally, the patient engages with interactive graphics to select from a menu of behavior change tips. To reach people with disabilities, the tool features closed captioning and keyboard navigation capabilities instead of a mouse or touchscreen. Extra Info, offered at the end of the tool, features peer-delivered educational videos and other resources.

Toward the end of the virtual clinic visit, PHC instantly generates two individually tailored handouts, a patient handout (Multimedia Appendices 4-5) and a summary handout for the provider, which the patient can elect to share or not. The patient can refer to the patient handout during the appointment and opt to send a PDF of it to their personal email account. As noted above, the handouts feature the patient's self-selected behavior change strategies and questions for their provider.

Challenges and Solutions to Designing and Developing PHC

Overview

This section of the tutorial describes guidance on key solutions and challenges we faced in developing PHC around three topical areas related to user engagement, integration of intervention assets, and building the app to deliver complex tailored content (Table 1).

Table 1. Overview of challenges to and solutions for designing and developing Positive Health Check.

Challenge	Solution
Topic 1. Enhancing user engagement through video design and production	
Leverage the medium of video to maximize user engagement	• Limit distractions and make it look real • Use entire video screen as a window into a virtual experience • Streamline the scripts
Plan for filming by visualizing the flow of scenes for the entire intervention	Develop storyboards that depict a series of images portraying actors' positions on the screen along with their dialogue
Topic 2. Linking videos and other digital assets through script coding and flow diagrams	
Linking the 700 videos to intervention scenarios while filming to ensure users' tailored experience	Coding the shooting script to link to video clips
Create a tool to monitor filming on the set	Create a master flow diagram of algorithms for each possible Positive Health Check user experience
Topic 3. Building the Positive Health Check app	
Programming the intervention tool	
How to plan to program the integration of video with multiple digital assets	Use keyframes to design the video screen by integrating video and other digital assets
Test the integrity of the functional requirements used to design Positive Health Check	Use agile methods to gather rapid iterative feedback from a multidisciplinary team
Deliver and display on-screen tailored content in real time as user inputs data	Build intelligent back-end design to support the delivery and display of users' tailored content
Building a tool that performs efficiently on multiple devices	
Deliver a video-based intervention on mobile devices with limited resources, such as processor and memory (random access memory)	Use a single-page app that preloads common media/JavaScript files and avoid multiple page refreshes
Use autoplaying of video clips and audio prompts on mobile devices	Develop hybrid apps that use device-specific settings to autoplay video
Finding a streaming video solution	
Selection of platform to ensure seamless delivery of sharp video with indiscernible buffering	Use streaming video from the Centers for Disease Control and Prevention web server
Planning for dissemination	
Ensure eventual dissemination and national-level scale-up to HIV clinics	Use of open web platform

Topic 1: Enhancing User Engagement Through Video Design and Production

Overview

We recognized that several key components were needed to create a video tool that would bring about user engagement and lead patients to become more active in their own health. In this context, user engagement refers to "the micro-level of moment-to-moment engagement with the intervention" designed to promote behavior change to improve health outcomes [31].

Challenge: Leverage the Medium of Video to Maximize User Engagement

User engagement is a critical process to improve intervention effectiveness and behavior change [31]. Our challenge was to increase users' engagement by limiting distractions, an important factor that informed creative decisions when designing PHC. Consequently, the visuals, location, acting style, wardrobe, dialogue, and camera movement all had to look real, be inviting, and impart confidence in the messages being received, without being distracting to the user.

Solution: Limit Distractions and Make It Look Real

We approached this challenge to limit distractions by enhancing the "realness" of the virtual doctor visit. Although we considered using real-life doctors for the PHC, we ultimately decided to use actors because they are trained to deliver scripts verbatim from teleprompters in front of stage lights and crew. This ensured consistency of the message and delivery across the four virtual doctors embedded in the intervention. Actors appeared in clothing, hair, and makeup that would not reflect a specific decade and soon appear outdated. To mimic what a conversation might be like in a real doctor visit, instead of presenting all of the risk behavior assessment tailoring questions at the beginning of the tool, we embedded these questions throughout the experience.

Solution: Use Entire Video Screen as a Window Into a Virtual Experience

We used a "corner to corner" 16 × 9 video screen that most closely approximates human natural visual perception, thus using the entire screen of the devices that would deliver the intervention. We used a video game structure to model high levels of engagement and immersion, treating each PHC screen depicted in the intervention as a "window" into a virtual experience. Finally, we overlaid graphics on top of the video so the user could still see the video image. This allows users to process the onscreen information but remain engaged and be rooted in the virtual doctor visit.

We also limited visual distractions by ensuring that the users' focus stayed on the actor/doctor and their health messages, rather than on something in the background of the set. We accomplished this by filming in full high definition (1920 × 1080) on a Sony FS700 because the camera's sensor size was large enough to provide a shallower depth of field, which would allow the background to be slightly blurred.

Solution: Streamline the Scripts

Although the original scripts were developed collaboratively with end users and HIV providers [29], the creative director continued to translate these information-dense scripts into lines the actors could easily convey to PHC users and imbued them with more natural, conversational language. This resulted in a streamlined script to ensure end users would be more engaged and attentive.

Challenge: Plan for Filming by Visualizing the Flow of Scenes for the Entire Intervention

User inputs resulted in the delivery of tailored video clips and layered multimedia digital assets. Achieving these tailored responses required filming 168 separate video scenes for each of the four actors playing virtual doctors, which eventually would be programmed to play in multiple combinations. The challenge was to correctly plan for filming by visualizing the flow of scenes for each patient as if they were all one seamless interaction.

Solution: Develop Storyboards That Depict a Series of Images Portraying the Actors' Position on the Screen Along With Their Dialogue

To break down the complexity of PHC, we used storyboards to help visualize the look, feel, and flow of the tool. We created storyboards for every major PHC scenario. The storyboards were created by shooting still images of the location, then overlaying representations of the actor and graphics that would appear on screen. Those images were then paired with the relevant scripting (Multimedia Appendices 6-7).

Topic 2: Linking Videos and Other Digital Assets Through Script Coding and Flow Diagrams

Overview

Video is the basic medium of PHC, along with the other digital assets described above: interactive and tailored graphics, the repurposed Safe in the City animated video, closed captions, and user-generated handout templates. To make the intervention highly tailored and responsive to user input, we needed to connect scripted messages to a database of videos and digital assets so that when prompted by user input, the tool would correctly display these digital assets in congruity (or in sync) with the video.

Challenge: Linking the 700 Videos to Intervention Scenarios While Filming to Ensure Users' Tailored Experience

PHC features 168 video clips for each of the four virtual doctors, and 28 clips common across doctors. It quickly became clear that all clips needed to be coded so we could efficiently coordinate the project. The challenge was to track the filming of all scenarios and ensure the correct use of all video clips according to programming algorithms. We needed a tool to coordinate the filming of all the clips and to ensure that when a user made selections while engaging with the intervention, PHC seamlessly delivered the appropriate video clip.

Solution: Coding the Shooting Script to Link to Video Clips

The shooting script links video clips with tailored messages and user inputs—such as when the user selects topics for behavior change tips—and provides direction to the video director on how to compose each shot. The shooting script also described any graphics that would appear on the screen. Coding the videos based on their respective PHC intervention modules and creating a video catalog (codebook) from the shooting script standardized a naming convention for the 700 video clips. Consequently, the shooting script became the ultimate tool to link PHC videos to all intervention scenarios and user inputs.

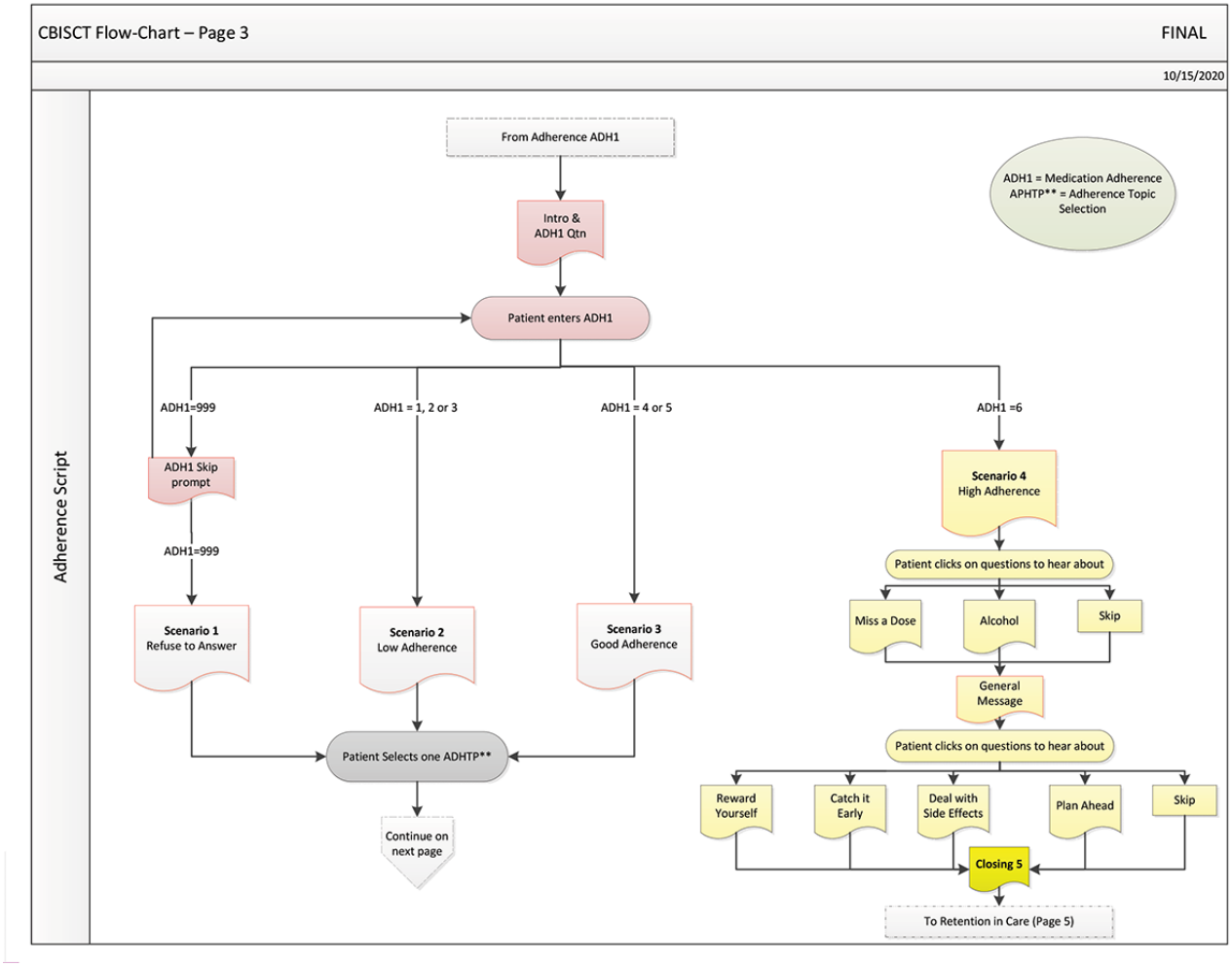
Challenge: Create a Tool to Monitor Filming on the Set

PHC was designed as a highly tailored intervention with hundreds of different possible user pathways depending on user input. The challenge was to create a tool to plan for and monitor filming so that the team could see the logic and interconnection between all of the scenes and understand how the composition of each frame would inform the composition of subsequent video frames.

Solution: Create a Master Flow Diagram of Algorithms for Each Possible PHC User Experience

The solution to this challenge was to create the PHC flow diagram to show all the different tailoring paths a patient could take through the intervention. The flow diagram was used by the creative team to monitor the filming of the intervention and by the intervention programmer to guide the development of the algorithms underlying the intervention structure. This ensured that each scene would flow smoothly and be perceived as a natural progression in a conversation between a doctor and a patient. Figure 1 shows an example flow diagram for the adherence module.

Figure 1. Positive Health Check flow diagram: adherence module pathway.



Topic 3: Building the PHC App

Overview

The PHC architecture achieves the following: (1) an intelligent, "back-end" design that refers to the underlying software framework (the app and the database) to provide tailored content based on user input, (2) a user-friendly interface that performs efficiently on multiple devices, such as computers, laptops, and tablets, and (3) a streaming video delivery solution that delivers high-quality video content.

The build-out started with the abovementioned coded shooting scripts and the tailoring flow diagram because these tools had applications for both filming and programming and were crucial linkage tools between the design and development teams at this stage. Building PHC involved finalizing technical functional requirements, prototyping, coding and programming, and performing internal quality testing ("sprints"). We also developed and tested each line of closed captioning that was paired with all script lines to ensure the captioning met Section 508 compliance requirements. Below, we describe the challenges and solutions encountered when building PHC, including programming the intervention tool, building a tool that performs efficiently on multiple devices, finding a streaming video solution, and planning for dissemination.

Programming the Intervention Tool

Challenge: How to Plan to Program the Integration of Video With Multiple Digital Assets

Each PHC module is constructed as a sequence of scenes appearing one after another, posing questions and imparting health messages. Each scene incorporates videos; input elements (such as check boxes) to collect user responses; display overlays to show topics and questions for patients to consider; and audio prompts to encourage patients to reengage during moments of inactivity, complete all questions, or avoid an early exit from the intervention. All the visual, HTML, and audio elements were choreographed in a scene to appear and animate on cue to provide the user with a visually rich and engaging experience. The challenge was how to create a tool to plan for, visualize, and document the placement of all these elements so that the videographer would frame each shot to give programmers the needed space for each element on the screen.

Solution: Use Keyframes to Design the Video Screen by Integrating Video and Other Digital Assets

In film and video production, keyframes are pictures that define the timing and the starting and end points of transitions. Specifically, PHC keyframes in congruity with the scripts were the solution for defining the positioning and movement of the actors and displaying overlay images, check boxes, navigation buttons, and closed captioning. Keyframes proved to be a

reliable tool for the graphic designers to communicate the intended look and user interface details to the developer who programmed the intervention. Multiple keyframes were created for each scene to show how visual elements would appear in the video at a specific point in time along with the narrative delivered by a doctor or nurse. Multimedia Appendix 8 shows an example of a keyframe for the adherence module, showing an actor/doctor in relation to the potential user and the graphic depicting choices of behavioral strategies designed to support HIV medication adherence, as well as navigation buttons.

Challenge: Test the Integrity of the Functional Requirements Used to Design PHC

The design of PHC had 200 functional requirements to guide the programming and delivery of 700 video clips to users. These functional requirements addressed the needs for specific intervention scenes, including the visual elements, data collected from or displayed to the patient, script to be conveyed through video, and all the probable user actions with corresponding responses from the system. Functional requirements also included rules to determine the next scene based on the user's response to the current scene and/or all of the previous scenes. The challenge was to correctly develop, implement, and troubleshoot the functional requirements that would guide PHC programming and specify the interaction between user inputs and PHC outputs.

Solution: Use Agile Methods to Gather Rapid Iterative Feedback From the Multidisciplinary Team and End Users

To address the challenge to verify and test the 200 functional requirements, we conducted numerous rapid and iterative feedback sessions or "sprints" in the application of "agile" methodology via collaborative, cross-functional, multidisciplinary teams [37,38]. As noted earlier, we also consulted end users, including patients and clinicians, across all phases of development [29]. This also entailed user testing of the prototypes and features during development. The solution was to rapidly develop, evaluate, and improve the PHC intervention tool to optimize and ready it for piloting. For example, in two sprints sessions, we resolved video buffering issues. To avoid buffering, we aimed to reduce video file size without increasing graininess or otherwise impacting the quality of the videos. After viewing several different resolutions, we selected an ideal file size to stream over a standard internet connection that would maintain video quality and eliminate buffering.

Challenge: Deliver and Display On-Screen Tailored Content in Real Time as the User Inputs Data

The individually tailored content distinguishes PHC from other digital interventions such as those delivered through text messaging. As noted previously, user engagement is critical to intervention effectiveness and behavior change. The challenge was to figure out how to design PHC to deliver and display tailored messages in real time along with all other tailored content from a large database of videos and other digital assets.

Solution: Build Intelligent Back-end Design to Support the Delivery and Display of Users' Tailored Content

As noted earlier, the PHC back-end design refers to the underlying software and app, as well as a database with data sets that define the relationship between the scenes and videos, audio prompts, questions, and tips used to tailor PHC content to each user. This back-end design supports the delivery of individually tailored health messages based on user input. To achieve this end, our solution was to build an intelligent intervention engine in the back end that connects all app assets defined in data sets for a given scene and is fully data driven (ie, responsive to user input). It was built on the Microsoft ASP.NET MVC platform and an SQL Server database. The individual modules (such as adherence or retention in care) and the related scenes with video messages and graphical overlays are depicted in various configuration tables that reside in the SQL Server database.

Building a Tool That Performs Efficiently on Multiple Devices

Challenge: Deliver a Video-Based Intervention on Mobile Devices With Limited Resources, Such As Processor and Memory (Random Access Memory)

PHC was required to perform efficiently on mobile devices in spite of the challenge presented by the fact that, unlike desktop or laptop computers, mobile devices have limited memory and processing capabilities. These limiting features accommodate the small size of mobile devices and prevent overheating in the absence of fans and ventilation. However, the drawback is that these devices tend to crash when loading webpages with multiple media and code files. This is especially true when multiple full-page refreshes are needed during typical use of a website. In these mobile devices, memory and processing power are limited but valuable resources that need to be used efficiently to make intervention pages load and perform optimally.

Solution: Use a Single-Page App That Preloads Common Media/JavaScript Files and Avoid Multiple Page Refreshes

To address this challenge, PHC was developed as a single-page app that loads a webpage only once, at the beginning of the session; after that, the webpage content is manipulated through JavaScript programming. Additionally, the loading of required content is accomplished through asynchronous data transfers. The single-page app allows PHC to preload the video and JavaScript files, which are used across all or most of the intervention scenes during the initial page load. This approach also avoids full-page refreshes when transitioning from one scene to the next; instead, the scene-specific data, display overlays, and scripts are loaded asynchronously from the back end of the app based on the user inputs and intervention scenario into a screen area defined in the single-page app. The preloading of common videos and scripts into the tool significantly reduces load time for individual scenes. Additionally, the asynchronous loading of content into a single-page app decreases potential crashes by avoiding multiple-page refreshes involving videos, images, and JavaScript files.

Challenge: Use Autoplaying of Video Clips and Audio Prompts on Mobile Devices

Another key requirement for PHC is to provide an engaging intervention user experience by transitioning from one scene to the next without the need for the user to manually start and stop videos. All videos need to play automatically. PHC also has audio prompts to keep the user focused by encouraging them to answer questions and alerting them to any periods of inactivity. These audio prompts are programmed to automatically play in sync with all the video questions and messages, which worked flawlessly in computer web browsers. However, a problem occurred with the autoplaying of videos on Apple and Android tablets. These mobile operating system vendors restrict video autoplay within their web browsers, such as Safari or Chrome. These restrictions prevent unexpected user data charges, which can be particularly expensive for cellular internet connections. Consequently, the challenge was to figure out how to play PHC videos on mobile devices.

Solution: Develop Hybrid Apps That Use Device-Specific Settings to Autoplay Video

To address this issue, we developed hybrid apps for Apple iPads and Android tablets. Hybrid apps are web apps wrapped inside native apps that are deployed in mobile app stores, such as Google Play or Apple Store. This hybrid approach allowed for the autoplay of PHC videos on these devices. In addition, the PHC hybrid app will allow for continual enhancements to PHC and modification of the user experience. Most importantly, by using this solution, we can deploy changes to PHC through the web without needing to update and deploy mobile apps each time there is an update to PHC.

Finding a Streaming Video Solution

Challenge: Selection of a Platform to Ensure Seamless Delivery of Sharp Video With Indiscernible Buffering

This challenge involved identifying a server that would efficiently deliver video content to PHC users. Cost, flexibility, data security, and streaming speed were important considerations for the final solution. YouTube provides a low-cost streaming option but offers little flexibility when it comes to controlling and displaying content. Videos hosted by YouTube, for example, cannot be played directly through the PHC app's HTML5 player, but instead must use Google's YouTube player. Additionally, the YouTube player always includes Google's advertisements and usage tracking code. As an alternate streaming solution, we considered Akamai's cloud delivery platform, which uses caching to rapidly stream content from servers globally. However, Akamai pricing was prohibitive.

Solution: Use Streaming Video From the CDC Web Server

After researching the above potential solutions to this challenge, we decided to implement a streaming video solution by hosting video content on a CDC internal web server. This decision was confirmed after evaluating video performance on delivering video content rapidly and securely using load testing with many concurrent users. The videos were fine-tuned to be of good quality yet lower file size to support delivery over standard internet connections. Although this approach is adequate for use within clinical settings, it may not scale well as the number of concurrent users increases. We recommend the use of cloud-based content delivery networks to stream videos for wider dissemination.

Planning for Dissemination

Challenge: Ensure Eventual Dissemination and National-Level Scale-up to HIV Clinics

We designed and developed PHC to create conditions for greater population impact, and decisions were made with the goal of disseminating PHC to HIV clinics, if it was found to be effective. We aimed to build an intervention that could be feasibly implemented in busy HIV clinic environments and would be relevant to clinic providers and patients with HIV engaged in clinical care. The challenge was to design an intervention tool that will be available and accessible on multiple commonly used devices.

Solution: Use of Open Web Platform

We achieved a solution for possible future compatibility by developing PHC using tools and technologies supporting the World Wide Web Consortium's open web standards. Apps using the open web platform are scalable, cross-platform compliant, and supported by multiple browsers without sacrificing user experience [39]. Standards-based systems are nonproprietary and include, for example, HTML5 and JavaScript. The front end of PHC is built on these open web standards–based tools. All the graphical interfaces are developed using HTML enriched for an interactive mobile experience using well-designed graphical images. This is preferable to creating a solution using third-party commercial off-the-shelf tools, such as Adobe Flash (Adobe), and enables PHC to run on any standard web browser without the need to install additional software tools.

Conclusion

Video-based digital health interventions will continue to play an important role in the future of HIV care and other clinical health practices. This is, in part, due to the ever-increasing popularity of video as a way to consume information across all sectors of society and multiple disciplines, including medicine [40,41]. Notably, the condom use animated video clip from Safe in the City (now repurposed for PHC) has received 980,000 views on YouTube. However, facilitating the adoption of an HIV video intervention in HIV clinical settings is still a work in progress. It remains a challenge to build individually tailored interventions that engage users, keep interventions up to date, and implement and scale up interventions in busy clinic workflows. We chose to design a tailored video-based intervention with the hope that it could meet these challenges and benefit the health of patients with HIV engaged in clinical care and improve clinics' performance metrics around viral suppression and retention in care. We believe that the return on investment of developing this highly interactive and tailored intervention can be considered beneficial if the intervention is shown to be effective and scaled up and implemented at a low cost by HIV clinics. We are currently conducting a pragmatic type 1 hybrid trial [42] to test the effectiveness of PHC in improving health outcomes for patients with HIV engaged in clinical care, as well as to test the feasibility of implementing

PHC in HIV clinics. The study also includes collecting data on the costs associated with implementing PHC in clinical settings [43].

Our experience in designing and developing PHC presented unique challenges because of the extensive use of a large database of videos individually tailored to each user, and also revealed other challenges that are well known to investigators who seek to develop digital interventions [44]. We believe that researchers and practitioners who seek to support and change health behavior using digital technology can benefit from our learnings, because the speed by which digital health interventions capitalize on technological advances far outpaces the development, testing, and deployment of health behavior change interventions. The tutorial presented here on the complex challenges and solutions involved in designing and developing PHC aims to support researchers in HIV and beyond who wish to keep pace with approaches to and the technology behind video-based digital interventions [37,38].

Acknowledgments

Support for this paper was provided by a cooperative agreement from the Centers for Disease Control and Prevention (1U18PS004967-01; principal investigator: MAL). The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the US Centers for Disease Control and Prevention.

We acknowledge the contributions of other PHC team members who are not coauthors on this manuscript: Alexa Ortiz, Robert Furberg, and Kate Ferriola-Bruckenstein from RTI International, who assisted with helping ensure the intervention was developed using human-centered design principles and would fit into a busy clinic environment; and Cynthia Lyles, Aisha Wilkes, Arin Freeman, and Gary Marks from the Centers for Disease Control and Prevention, and Christine O'Daniels from Carter Consulting Inc, for their contribution to the conception of the intervention and input on many design aspects of the intervention. We also thank the following from Emergint Technologies Inc, who helped with development, graphics, user interface development, and architecture: Khongbeng Lue Xiong, Frank Robinson, and Rishi Tarar.

Authors' Contributions

Each coauthor contributed to the conception and design of the manuscript, drafted manuscript sections, and contributed to critical revisions. CH is the guarantor.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Positive Health Check: Extra Info.
[DOCX File , 32 KB - mhealth_v9i3e21128_app1.docx]

Multimedia Appendix 2
Positive Health Check: Extra Info landing page.
[PNG File , 915 KB - mhealth_v9i3e21128_app2.png]

Multimedia Appendix 3
Positive Health Check: Your Sexual Relationships.
[PNG File , 695 KB - mhealth_v9i3e21128_app3.png]

Multimedia Appendix 4
Positive Health Check patient handout (image 1).
[PNG File , 204 KB - mhealth_v9i3e21128_app4.png]

Multimedia Appendix 5
Positive Health Check patient handout (image 2).
[PNG File , 115 KB - mhealth_v9i3e21128_app5.png]

Multimedia Appendix 6
Excerpt of Positive Health Check storyboard: opening script (image 1).
[PNG File , 481 KB - mhealth_v9i3e21128_app6.png]

Multimedia Appendix 7

Excerpt of Positive Health Check storyboard: opening script (image 2).
[PNG File , 556 KB - mhealth_v9i3e21128_app7.png]

Multimedia Appendix 8
Positive Health Check keyframe for the adherence module (image 2).
[PNG File , 391 KB - mhealth_v9i3e21128_app8.png]

References

1. Four important trends in digital media. 2017 Jul 10. URL: https://www.hiv.gov/blog/four-important-trends-digital-media [accessed 2018-08-23]
2. Blake KD, Thai C, Falisi A, Chou WS, Oh A, Jackson D, et al. Video-Based Interventions for Cancer Control: A Systematic Review. Health Educ Behav 2020 Apr;47(2):249-257. [doi: 10.1177/1090198119887210] [Medline: 31701780]
3. Oosterveen E, Tzelepis F, Ashton L, Hutchesson MJ. A systematic review of eHealth behavioral interventions targeting smoking, nutrition, alcohol, physical activity and/or obesity for young adults. Prev Med 2017 Jun;99:197-206. [doi: 10.1016/j.ypmed.2017.01.009] [Medline: 28130046]
4. Stephenson A, McDonough SM, Murphy MH, Nugent CD, Mair JL. Using computer, mobile and wearable technology enhanced interventions to reduce sedentary behaviour: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2017 Aug 11;14(1):105 [FREE Full text] [doi: 10.1186/s12966-017-0561-4] [Medline: 28800736]
5. Fagherazzi G, Ravaud P. Digital diabetes: Perspectives for diabetes prevention, management and research. Diabetes Metab 2019 Sep;45(4):322-329. [doi: 10.1016/j.diabet.2018.08.012] [Medline: 30243616]
6. Centers for Disease Control and Prevention. High-impact HIV prevention: CDC's approach to reducing HIV infections in the United States.: Center for Disease Control and Prevention, National Center for HIV/AIDS, Viral Hepatitis, STD and TB Prevention, Division of HIV/AIDS Prevention; 2011. URL: https://www.cdc.gov/hiv/policies/hip/hip.html [accessed 2017-05-09] [WebCite Cache ID https://www.cdc.gov/hiv/policies/hip/hip.html]
7. Greenberg AE, Gordon CM, Purcell DW. Promotion of Research on the HIV Continuum of Care in the United States: The CFAR HIV Continuum of Care/ECHPP Working Group. J Acquir Immune Defic Syndr 2017 Feb 01;74 Suppl 2:S75-S80 [FREE Full text] [doi: 10.1097/QAI.0000000000001243] [Medline: 28079716]
8. Daher J, Vijh R, Linthwaite B, Dave S, Kim J, Dheda K, et al. Do digital innovations for HIV and sexually transmitted infections work? Results from a systematic review (1996-2017). BMJ Open 2017 Nov 03;7(11):e017604 [FREE Full text] [doi: 10.1136/bmjopen-2017-017604] [Medline: 29101138]
9. Muessig KE, Nekkanti M, Bauermeister J, Bull S, Hightow-Weidman LB. A Systematic Review of Recent Smartphone, Internet and Web 2.0 Interventions to Address the HIV Continuum of Care. Curr HIV/AIDS Rep 2015 Jan 28. [doi: 10.1007/s11904-014-0239-3] [Medline: 25626718]
10. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
11. Goldenberg T, McDougal SJ, Sullivan PS, Stekler JD, Stephenson R. Preferences for a Mobile HIV Prevention App for Men Who Have Sex With Men. JMIR Mhealth Uhealth 2014;2(4):e47 [FREE Full text] [doi: 10.2196/mhealth.3745] [Medline: 25355249]
12. Goldenberg T, McDougal SJ, Sullivan PS, Stekler JD, Stephenson R. Building a Mobile HIV Prevention App for Men Who Have Sex With Men: An Iterative and Community-Driven Process. JMIR Public Health Surveill 2015;1(2):e18 [FREE Full text] [doi: 10.2196/publichealth.4449] [Medline: 27227136]
13. Horvath KJ, Alemu D, Danh T, Baker JV, Carrico AW. Creating Effective Mobile Phone Apps to Optimize Antiretroviral Therapy Adherence: Perspectives From Stimulant-Using HIV-Positive Men Who Have Sex With Men. JMIR Mhealth Uhealth 2016 Apr 15;4(2):e48 [FREE Full text] [doi: 10.2196/mhealth.5287] [Medline: 27084049]
14. Mitchell JW, Torres MB, Joe J, Danh T, Gass B, Horvath KJ. Formative Work to Develop a Tailored HIV Testing Smartphone App for Diverse, At-Risk, HIV-Negative Men Who Have Sex With Men: A Focus Group Study. JMIR Mhealth Uhealth 2016 Nov 16;4(4):e128 [FREE Full text] [doi: 10.2196/mhealth.6178] [Medline: 27852558]
15. 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 and Behavior 2017 Oct 05;22:43-63. [doi: 10.1007/s10461-017-1923-2]
16. Cao B, Gupta S, Wang J, Hightow-Weidman LB, Muessig KE, Tang W, et al. Social Media Interventions to Promote HIV Testing, Linkage, Adherence, and Retention: Systematic Review and Meta-Analysis. J Med Internet Res 2017 Nov 24;19(11):e394 [FREE Full text] [doi: 10.2196/jmir.7997] [Medline: 29175811]
17. O'Donnell CR, O'Donnell L, San Doval A, Duran R, Labes K. Reductions in STD infections subsequent to an STD clinic visit. Using video-based patient education to supplement provider interactions. Sex Transm Dis 1998 Mar;25(3):161-168. [doi: 10.1097/00007435-199803000-00010] [Medline: 9524995]

18. Harshbarger CL, O'Donnell LN, Warner L, Margolis AD, Richardson DB, Novey SR, et al. Safe in the city: effective prevention interventions for human immunodeficiency virus and sexually transmitted infections. Am J Prev Med 2012 May;42(5):468-472. [doi: 10.1016/j.amepre.2012.01.029] [Medline: 22516486]
19. O'Donnell LN, Doval AS, Duran R, O'Donnell C. Video-based sexually transmitted disease patient education: its impact on condom acquisition. Am J Public Health 1995 Jun;85(6):817-822. [doi: 10.2105/ajph.85.6.817] [Medline: 7762716]
20. Warner L, Klausner JD, Rietmeijer CA, Malotte CK, O'Donnell L, Margolis AD, Safe in the City Study Group. Effect of a brief video intervention on incident infection among patients attending sexually transmitted disease clinics. PLoS Med 2008 Jun 24;5(6):e135 [FREE Full text] [doi: 10.1371/journal.pmed.0050135] [Medline: 18578564]
21. McBride CM, Emmons KM, Lipkus IM. Understanding the potential of teachable moments: the case of smoking cessation. Health Educ Res 2003 Apr;18(2):156-170. [doi: 10.1093/her/18.2.156] [Medline: 12729175]
22. Horvath KJ, Ecklund AM, Hunt SL, Nelson TF, Toomey TL. Developing Internet-based health interventions: a guide for public health researchers and practitioners. J Med Internet Res 2015;17(1):e28 [FREE Full text] [doi: 10.2196/jmir.3770] [Medline: 25650702]
23. Fisher JD, Fisher WA. Changing AIDS-risk behavior. Psychol Bull 1992 May;111(3):455-474. [doi: 10.1037/0033-2909.111.3.455] [Medline: 1594721]
24. Miller WR, Rollnick S. Motivational interviewing: helping people change (third edition). New York, NY: Guildford Press; 2012.
25. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
26. Fisher JD, Amico KR, Fisher WA, Cornman DH, Shuper PA, Trayling C, et al. Computer-based intervention in HIV clinical care setting improves antiretroviral adherence: the LifeWindows Project. AIDS Behav 2011 Nov;15(8):1635-1646. [doi: 10.1007/s10461-011-9926-x] [Medline: 21452051]
27. Kurth AE, Spielberg F, Cleland CM, Lambdin B, Bangsberg DR, Frick PA, et al. Computerized counseling reduces HIV-1 viral load and sexual transmission risk: findings from a randomized controlled trial. J Acquir Immune Defic Syndr 2014 Apr 15;65(5):611-620 [FREE Full text] [doi: 10.1097/QAI.0000000000000100] [Medline: 24384803]
28. Gilbert P, Ciccarone D, Gansky SA, Bangsberg DR, Clanon K, McPhee SJ, et al. Interactive. PLoS One 2008 Apr 23;3(4):e1988 [FREE Full text] [doi: 10.1371/journal.pone.0001988] [Medline: 18431475]
29. Harshbarger C, Taylor O, Uhrig JD, Lewis MA. Positive health check: developing a web-based video counseling tool for HIV primary care clinics. Journal of Communication in Healthcare 2017 Jul;10(2):70-77. [doi: 10.1080/17538068.2017.1341189]
30. Latimer N. How technology is shaping patient education. Cerner. 2018. URL: https://www.cerner.com/blog/how-technology-is-shaping-patient-education [accessed 2021-03-10]
31. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and Promoting Effective Engagement With Digital Behavior Change Interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
32. Kelley JM, Kraft-Todd G, Schapira L, Kossowsky J, Riess H. The influence of the patient-clinician relationship on healthcare outcomes: a systematic review and meta-analysis of randomized controlled trials. PLoS One 2014;9(4):e94207 [FREE Full text] [doi: 10.1371/journal.pone.0094207] [Medline: 24718585]
33. Lokhorst AM, Werner C, Staats H, van Dijk E, Gale JL. Commitment and Behavior Change. Environment and Behavior 2011 Jun 09;45(1):3-34. [doi: 10.1177/0013916511411477]
34. The Centers for Disease Control and Prevention. Safe in the City: How to Use a Condom. URL: https://www.youtube.com/watch?v=tcpfZKvOFZ4 [accessed 2021-03-11]
35. Barnett TJ, Jain S, Andra U, Khurana T. Cisco Visual Networking Index (VNI) Complete Forecast Update, 2017-2022. 2018. URL: https://www.cisco.com/c/dam/m/en_us/network-intelligence/service-provider/digital-transformation/knowledge-network-webinars/pdfs/1211_BUSINESS_SERVICES_CKN_PDF.pdf [accessed 2021-03-11]
36. Centers for Disease Control and Prevention. Positive Health Check: Adherence module excerpt. 2015. URL: https://www.youtube.com/watch?v=iIyQ7O_wxvA [accessed 2021-03-11]
37. Hekler EB, Klasnja P, Harlow J. Agile science. In: Gellman M, Turner JR, editors. Encyclopedia of Behavioral Medicine. New York, NY: Springer; 2013.
38. Hekler EB, Klasnja P, Riley WT, Buman MP, Huberty J, Rivera DE, et al. Agile science: creating useful products for behavior change in the real world. Transl Behav Med 2016 Jun;6(2):317-328 [FREE Full text] [doi: 10.1007/s13142-016-0395-7] [Medline: 27357001]
39. Xanthopoulos S, Xinogalos S. A comparative analysis of cross-platform development approaches for mobile applications. In: BCI '13: Proceedings of the 6th Balkan Conference in Informatics. New York, NY: ACM Press; 2013 Presented at: 6th Balkan Conference in Informatics; September 13; Thessaloniki, Greece p. 213-220. [doi: 10.1145/2490257.2490292]
40. Tackett S, Slinn K, Marshall T, Gaglani S, Waldman V, Desai R. Medical Education Videos for the World: An Analysis of Viewing Patterns for a YouTube Channel. Acad Med 2018 Aug;93(8):1150-1156. [doi: 10.1097/ACM.0000000000002118] [Medline: 29298180]

41. Topps D, Helmer J, Ellaway R. YouTube as a platform for publishing clinical skills training videos. Acad Med 2013 Feb;88(2):192-197. [doi: 10.1097/ACM.0b013e31827c5352] [Medline: 23269305]
42. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care 2012 Mar;50(3):217-226 [FREE Full text] [doi: 10.1097/MLR.0b013e3182408812] [Medline: 22310560]
43. Lewis MA, Harshbarger C, Bann C, Burrus O, Peinado S, Garner BR, Positive Health Check Study Team. Positive Health Check evaluation: A type 1 hybrid design randomized trial to decrease HIV viral loads in patients seen in HIV primary care. Contemp Clin Trials 2020 Sep;96:1-9 [FREE Full text] [doi: 10.1016/j.cct.2020.106097] [Medline: 32738408]
44. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and Evaluating Digital Interventions to Promote Behavior Change in Health and Health Care: Recommendations Resulting From an International Workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]

Abbreviations

ART: antiretroviral therapy
CDC: Centers for Disease Control and Prevention
PHC: Positive Health Check
STD: sexually transmitted disease

Edited by L Buis; submitted 05.06.20; peer-reviewed by C Thai, W Tang; comments to author 20.07.20; revised version received 28.10.20; accepted 15.02.21; published 23.03.21.

Please cite as:
Harshbarger C, Burrus O, Rangarajan S, Bollenbacher J, Zulkiewicz B, Verma R, Galindo CA, Lewis MA
Challenges of and Solutions for Developing Tailored Video Interventions That Integrate Multiple Digital Assets to Promote Engagement and Improve Health Outcomes: Tutorial
JMIR Mhealth Uhealth 2021;9(3):e21128
URL: https://mhealth.jmir.org/2021/3/e21128
doi:10.2196/21128
PMID:33755025

Original Paper

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

David Joseph Muggeridge[1*], BSc, PhD; Kirsty Hickson[2], PgDip; Aimie Victoria Davies[3], BA, BSc; Oonagh M Giggins[4], BSc, MSc, PhD; Ian L Megson[2], BSc, PhD; Trish Gorely[3], BPhEd, MEd, PhD; Daniel R Crabtree[2*], BSc, MSc, PhD

[1]Edinburgh Napier University, Edinburgh, United Kingdom

[2]Division of Biomedical Sciences, University of the Highlands and Islands, Inverness, United Kingdom

[3]Department of Nursing and Midwifery, University of the Highlands and Islands, Inverness, United Kingdom

[4]NetwellCASALA, Dundalk Institute of Technology, Dundalk, Ireland

[*]these authors contributed equally

Corresponding Author:
David Joseph Muggeridge, BSc, PhD
Edinburgh Napier University
Sighthill Campus
Edinburgh, EH11 4BN
United Kingdom
Phone: 44 7979086243
Email: d.muggeridge@napier.ac.uk

Abstract

Background: Accurate, continuous heart rate measurements are important for health assessment, physical activity, and sporting performance, and the integration of heart rate measurements into wearable devices has extended its accessibility. Although the use of photoplethysmography technology is not new, the available data relating to the validity of measurement are limited, and the range of activities being performed is often restricted to one exercise domain and/or limited intensities.

Objective: The primary objective of this study was to assess the validity of the Polar OH1 and Fitbit Charge 3 devices for measuring heart rate during rest, light, moderate, vigorous, and sprint-type exercise.

Methods: A total of 20 healthy adults (9 female; height: mean 1.73 [SD 0.1] m; body mass: mean 71.6 [SD 11.0] kg; and age: mean 40 [SD 10] years) volunteered and provided written informed consent to participate in the study consisting of 2 trials. Trial 1 was split into 3 components: 15-minute sedentary activities, 10-minute cycling on a bicycle ergometer, and incremental exercise test to exhaustion on a motorized treadmill (18-42 minutes). Trial 2 was split into 2 components: 4 × 15-second maximal sprints on a cycle ergometer and 4 × 30- to 50-m sprints on a nonmotorized resistance treadmill. Data from the 3 devices were time-aligned, and the validity of Polar OH1 and Fitbit Charge 3 was assessed against Polar H10 (criterion device). Validity was evaluated using the Bland and Altman analysis, Pearson moment correlation coefficient, and mean absolute percentage error.

Results: Overall, there was a very good correlation between the Polar OH1 and Polar H10 devices (r=0.95), with a mean bias of −1 beats·min^{-1} and limits of agreement of −20 to 19 beats·min^{-1}. The Fitbit Charge 3 device underestimated heart rate by 7 beats·min^{-1} compared with Polar H10, with a limit of agreement of −46 to 33 beats·min^{-1} and poor correlation (r=0.8). The mean absolute percentage error for both devices was deemed acceptable (<5%). Polar OH1 performed well across each phase of trial 1; however, validity was worse for trial 2 activities. Fitbit Charge 3 performed well only during rest and nonsprint-based treadmill activities.

Conclusions: Compared with our criterion device, Polar OH1 was accurate at assessing heart rate, but the accuracy of Fitbit Charge 3 was generally poor. Polar OH1 performed worse during trial 2 compared with the activities in trial 1, and the validity of the Fitbit Charge 3 device was particularly poor during our cycling exercises.

(JMIR Mhealth Uhealth 2021;9(3):e25313) doi:10.2196/25313

KEYWORDS

heart rate; photoplethysmography; wearable electronic devices; validation study; exercise; mobile phone

Introduction

Background

Consumer wearables constitute an ever-evolving industry with applications across multiple sectors of society. One key demand for wearable technology is to monitor and use parameters associated with physical activity for sport performance, health, and well-being. For example, a recent systematic review and meta-analysis concluded that the utilization of a consumer-based wearable activity tracker, used either as the primary component of an intervention or as part of a broader physical activity intervention, has the potential to increase participation in physical activities [1].

The autonomic nervous system (ANS) is interlinked with many physiological systems, and heart rate (HR) measures are considered surrogate markers of ANS status [2,3]. Exercise training stimulates ANS status changes, and HR measures therefore reflect physiological responses related to training adaptations [4]. HR monitoring enables noninvasive, cost-effective, and continuous insights into exercise intensity [5]. This may be important for optimizing training responses and for safety considerations (eg, restricting exercise intensity because of contraindications associated with an elevated HR). Wrist-worn, optical HR sensors using photoplethysmography (PPG) technology are exciting, as they reduce the requirement for chest strap or electrocardiogram (ECG) device use. ECG devices are inconvenient, can cause discomfort, and may be time-consuming to use, potentially providing a barrier to participation and optimal training. Although the use of PPG technology is not new in devices, the range of activity types and intensities being performed in validation studies is often restricted. Despite the widely adopted use of PPG in wearable devices, the standards of accuracy and reliability outside of medical devices can be poor, technology frequently advances, and new products are continuously released. Furthermore, the algorithms adopted by companies to derive HR from raw PPG signals are constantly changing, thereby affecting the validity of wearable devices for measuring HR. It is therefore important that devices are continually subjected to scientific scrutiny for appropriate interpretation and advice to be provided.

Polar OH1 (Polar Electro Oy) and Fitbit Charge 3 (Fitbit Inc) are two of the latest available devices that use PPG technology. They differ in that the Fitbit device constitutes a watch worn on the wrist, whereas Polar OH1 may be worn on either the forearm or the temple and is a stand-alone optical HR sensor. It has previously been shown that Polar OH1 is accurate at measuring HR during moderate-intensity yoga compared with a Polar H7 chest strap (mean bias: −0.76 beats·min^{-1}; 95% limits of agreement [LoAs]: −3.83 to 5.35 beats·min^{-1}) [6] and during moderate-to-high intensity treadmill and spin bike exercise compared with ECG (forearm sensor aggregated data mean bias: 0.27 beats·min^{-1}; LoAs: −4.68 to 5.22 beats·min^{-1}) [7]. Conversely, although the validity of Fitbit Charge 2 has previously been assessed [8,9], we are unaware of any study assessing the validity of Fitbit Charge 3. Fitbit Charge 3 was released in 2018 as an upgrade to the subsequently discontinued Fitbit Charge 2; however, despite Fitbit Charge 2 demonstrating poor HR measurement accuracy at higher workloads [8,10] and during cycling [9,11], HR monitoring technology appears unchanged. Furthermore, limited data are available for a range of exercise types and intensities. Specifically, there are currently no studies that have assessed PPG HR validity during sprint interval exercise (SIE), defined as exercise bouts performed in an *all-out* manner or at an absolute intensity that exceeds the workload required to elicit maximal oxygen uptake, with each bout separated by a recovery interval [12]. This may be particularly pertinent given the current popularity of SIE and high-intensity interval exercise (HIIE), which is commonly defined as relatively intense bouts of exercise that elicit ≥80% of maximal HR, interspersed with recovery periods [12]. Several studies examining the validity of commercial wearable devices have reported an increase in HR measurement error as exercise intensity increases [8,9,13,14]. This has been attributed to increased motion artifacts caused by rapid arm swinging during running [13], although it has been suggested that sustained isometric muscle contractions when gripping handlebars during cycling may reduce contact between a wrist-worn device and the skin [15]. In addition, during the initiation of exercise, HR can change rapidly with a limited increase in forearm or wrist blood flow to the skin [16]. Therefore, it is important to assess the validity of HR measurements during different activities and intensities. To our knowledge, no study has investigated the accuracy of Polar OH1 or Fitbit Charge 3 across a range of activity types and intensities.

Objectives

Therefore, the objective of this study was to assess the validity of Polar OH1 and Fitbit Charge 3 for measuring HR during rest, cycling, walking, and running activities. Validity during cycling, walking, and running activities was assessed across a range of intensities, including light, moderate, vigorous, and uniquely sprint-type exercise.

Methods

Participants

A total of 20 healthy adults (9 female; height: mean 1.73 [SD 0.1] m; body mass: mean 71.6 [SD 11.0] kg; fat mass: mean 17.0% [SD 7.8%]; age: mean 40 [SD 10] years; and International Physical Activity Questionnaire-Short Form [IPAQ-SF] [17] Physical Activity Category: high=20, moderate=0, and low=0) volunteered and provided written informed consent to participate in the study. Inclusion criteria were age between 18 and 60 years and categorized as having moderate-to-high levels of physical activity (as determined by the IPAQ-SF). Exclusion criteria were categorized as having low levels of physical activity (as determined by the IPAQ-SF), a noncommunicable disease (eg, cardiovascular disease, cancer, and respiratory disease), musculoskeletal injury in the past 2 months, and illness in the previous 6 weeks. One female participant withdrew from the

study at visit 1 because of injury (n=19). All the remaining participants completed visit 1 (n=19); however, 1 female participant was unable to complete visit 2 because of illness (n=18), and 1 male participant was unable to complete the visit-2 treadmill sprints because of injury (treadmill sprints; n=17). Ethical approval was provided by the University Ethics Committee at the University of the Highlands and Islands (OL-ETH-SHE-1436). All procedures were conducted in accordance with the Declaration of Helsinki 1974 and its later amendments. Written informed consent was obtained from all volunteers before entering the study, and the participants could withdraw at any point.

Study Design

The study consisted of 2 visits to the Active Health Exercise Laboratory at the University of the Highlands and Islands, Inverness, which were conducted a minimum of 3 days apart. Participants were asked to refrain from intense physical activity (24 h), alcohol (12 h), caffeine (6 h), and food (3 h) before arrival at the laboratory for each visit. During visit 1, participants completed 15 minutes of sedentary activities, 10 minutes of cycling, and a treadmill protocol. For visit 2, each participant completed cycling and treadmill-based HIIE protocols. During each of the trials, participants' HRs were continuously monitored by a Polar H10 heart rate monitor (Polar Electro Oy; criterion measure), Polar OH1, and Fitbit Charge 3.

Devices

Polar H10 HR Monitor

Polar H10 was used as the criterion device. The HR sensor was attached to a Polar Pro heart rate strap placed over the sternum. Polar H10 live data were transmitted to a spiroergometry system (METALYZER 3B, CORTEX Biophysik GmbH), which recorded HR data at 1-second intervals. Polar H10 has previously been found to be valid when compared with ECG, with a correlation of r=0.997 [18].

Polar OH1

Polar OH1 was attached to an arm band and strapped securely to the nondominant forearm, according to the manufacturer's instructions. HR data were recorded at 1-second intervals using 6 light-emitting diode sensors and live transmitted via Bluetooth to a smartphone with the Polar Beat app (Polar Electro Oy). After completion of each visit, data were uploaded to the Polar Flow web service (Polar Electro Oy).

Fitbit Charge 3

Fitbit Charge 3 was attached to the nondominant wrist, 2-finger widths above the ulnar styloid process, following the manufacturer's instructions. According to the manufacturer, Fitbit Charge 3 uses *PurePulse* wrist HR technology to measure HR. Data were synced to an anonymized Fitbit account, and subsequently, the intraday second-by-second data were exported for each session using the opensource software *Pulse Watch* [19].

Study Procedures

Visit 1

Upon arrival at the laboratory, participants were briefed on the protocol before anthropometric variables were measured. Height, body mass, and body composition were measured with participants wearing light exercise clothing. Height was measured to the nearest 0.1 cm using a portable stadiometer (Model 213, Seca), body mass was measured to the nearest 0.1 kg using a floor scale (Model 875, Seca), and body composition was assessed using bioelectrical impedance (MC780MA, Tanita Corporation). The HR measuring devices were subsequently attached as described above.

The trial was split into 3 components. Component 1 consisted of 15 minutes of sedentary activities. The participants remained seated in a chair for the duration of component 1 and were instructed to keep their movement to a minimum. During the first 5 minutes, participants sat quietly before watching 5 minutes of a nature documentary. All participants watched the same 5-minute section of the documentary. Finally, the participants completed a cognitive task where they were provided with a choice of either a word search, crossword, or sudoku puzzle and were instructed to complete as much as possible within the time frame. Before component 2, a target HR range was calculated to determine 60%-85% of HR reserve target intensity using the equation:

$$\text{Target HR} = \text{Percentage target intensity} \times (\text{age-predicted max HR} - \text{resting HR}) + \text{resting HR}$$

Resting HR was calculated as an average of the final minute of sitting quietly, whereas the age-predicted maximum HR was calculated using the following formula by Gellish et al [20]:

$$\text{Age-predicted max HR} = 206.7 - (0.67 \times \text{age}).$$

Component 2 consisted of 10-minute cycling on a bicycle ergometer (Lode Corival). During the cycling tasks, each participant completed 5 minutes of light work at 50 W. After 5 minutes, the intensity was increased to substantially elevate HR to between 60% and 85% of the HR reserve. Between components 2 and 3, participants rested for a minimum of 10 minutes.

Component 3 consisted of an incremental exercise test on a motorized treadmill (Skillrun, Technogym). The treadmill test consisted of a range of intensities from light to high intensity, increasing at 3-minute intervals until volitional exhaustion. The initial phase of each test was continuous, beginning with walking at speeds of 4, 5, and 6 $km \cdot h^{-1}$ and then running at speeds of 8 and 10 $km \cdot h^{-1}$ at 1% gradient for 3 minutes. For those who were able to continue, a subsequent discontinuous phase of the test immediately followed the continuous phase. The discontinuous phase consisted of running for 3 minutes at 12, 14, 16, and 18 $km \cdot h^{-1}$ (or until volitional exhaustion), with each stage separated by 3 minutes of active recovery (walking at 4 $km \cdot h^{-1}$). The discontinuous phase was used to allow the participants to complete as many stages as possible. Participants were not required to complete all stages to be included in the analysis.

Visit 2

Upon arrival at visit 2, the devices were attached to the participants, as described in visit 1. Visit 2 was split into 2 components. Component 1 consisted of a 3-minute warm-up followed by 4 maximal sprints, each lasting 15 seconds and interspersed with 3-minute active recovery on a cycle ergometer (Wattbike Pro, Wattbike). The airbrake resistance was set at 1 during the warm-up and active recovery phases and was increased based on body mass, as per the manufacturer's guidelines (Multimedia Appendix 1), immediately before each sprint. Following the completion of a 3-minute active cooldown, participants were instructed to rest for 30 minutes before commencing component 2.

Component 2 was performed using the Technogym Skillrun treadmill's *parachute training mode*, which is a nonmotorized program and whose objectives are based on distance, not duration. In addition, treadmill belt resistance is adjusted by changing the *parachute size*, which ranges from extra small (lowest resistance setting) to extra large (highest resistance setting); however, the resistance that each parachute size pertains to has not been quantified by Technogym. For the purposes of this study, the distance to be covered was selected based on the final velocity of the incremental treadmill test at visit 1 (Multimedia Appendix 2), and the treadmill belt resistance was set based on body mass (Multimedia Appendix 3). Participants self-propelled the treadmill belt and were secured to the treadmill using a harness worn around their waist. Before starting the repeated sprint protocol, participants completed 400 m on the treadmill at a self-selected pace, with belt resistance at the lowest setting, to act as a familiarization and warm-up. They were then given the opportunity to perform the stretching exercises. The sprints consisted of 4 repetitions, each lasting approximately 15 seconds (range 13-18 s). It was anticipated that the duration of each sprint would last approximately 15 seconds to match the sprint cycling trial, as it is not possible to complete a parachute test on the treadmill in a time setting. Following each sprint, the treadmill was set to a motorized setting of 4 $km \cdot h^{-1}$ and 1% gradient for 3 minutes of active recovery.

Data Analysis

Data from the 3 devices were time-aligned and split into the following parts:

- Visit 1: rest, light cycling, vigorous cycling, and treadmill
- Visit 2: sprint cycling and sprint running

The validity of Polar OH1 and Fitbit Charge 3 was compared with the validity of Polar H10 (criterion device) for all data points and as an average HR for each segment. Data alignment and filtering were performed in R Studio using the packages dplyr and tidr. Before analysis, the normality of data was assessed using histograms and quantile-quantile plots. Validity was subsequently evaluated using the Bland and Altman [21] analysis. Secondary measures, including the Pearson moment correlation coefficient and mean absolute percentage error (MAPE), are also provided. The Bland and Altman [21] analysis was used to express agreement between the measured and predicted beats per minute, where 95% LoA was calculated as mean bias (1.96SD). We deemed a MAPE of 0%-5% to be within the acceptable limits, a commonly adopted approach [9,22]. Correlation coefficients were interpreted as very poor (r<0.69), poor (r=0.70-0.84), good (r=0.85-0.94), very good (r=0.95-0.994), and excellent (r>0.995). The MAPE was calculated using the following equation, which provided a general measurement error for the monitors:

$$MAPE=((monitor-criterion)/criterion100)$$

Results

Validity of HR Across All Data

Combined data across all activity types showed that Polar OH1 underestimated HR by 1 $beat \cdot min^{-1}$ (LoA: −20 to 19 $beats \cdot min^{-1}$) versus the Polar H10 device, and there was a very good correlation between the devices (r=0.957; Table 1). The Fitbit device underestimated the HR by 7 $beats \cdot min^{-1}$ (LoA: −46 to 33 $beats \cdot min^{-1}$; Table 1). Overall, there was a good correlation between Fitbit Charge 3 and Polar H10 (r=0.807), and the MAPE for both devices was deemed acceptable (Table 1). In addition, aggregated HR data for each activity domain revealed results similar to those of the unaveraged data analysis (Table 2).

Table 1. Validity of measuring heart rate with the Polar OH1 and Fitbit Charge 3 devices. Data are unaveraged across all data points (overall) and within each of the activity domains.

Device	Overall (n=35,639[a])	Rest (n=5448[a])	Cycling light (n=1903[a])	Cycling hard (n=1928[a])	Treadmill (n=14,489[a])	Sprint cycling (n=5516[a])	Sprint running (n=6355[a])
Polar H10							
Heart rate (beats·min^{-1}), mean (SD)	114 (33)	63 (10)	89 (12)	119 (16)	124 (31)	131 (25)	123 (19)
Polar OH1							
Heart rate (beats·min^{-1}), mean (SD)	113 (33)	63 (10)	89 (12)	118 (17)	124 (30)	125 (26)	125 (20)
Mean bias (beats·min^{-1})	−1	0	1	−1	0	−6	2
Limits of agreement (beats·min^{-1})	−20 to 19	−4 to 4	−5 to 7	−7 to 5	−9 to 8	−38 to 27	−27 to 31
MAPE[b] (%)	0.4	0.2	−0.8	0.6	0.2	3.9	−1.9
Correlation coefficient	0.957	0.974	0.983	0.985	0.990	0.794	0.722
95% CI of correlation coefficient	0.956 to 0.958	0.973 to 0.975	0.981 to 0.984	0.983 to 0.987	0.990 to 0.991	0.784 to 0.803	0.710 to 0.734
Fitbit Charge 3							
Heart rate (beats·min^{-1}), mean (SD)	107 (31)	62 (9)	81 (12)	98 (24)	123 (27)	111 (30)	114 (16)
Mean bias (beats·min^{-1})	−7	−1	−7	−21	−1	−20	−10
Limits of agreement (beats·min^{-1})	−46 to 33	−7 to 5	−38 to 23	−70 to 29	−30 to 28	−79 to 40	−49 to 30
MAPE (%)	−4.4	−1.4	−7.1	−16.4	0.7	−13.5	−6.3
Correlation coefficient	0.807	0.946	0.272	0.183	0.879	0.390	0.348
95% CI of correlation coefficient	0.804 to 0.811	0.943 to 0.949	0.230 to 0.328	0.127 to 0.238	0.875 to 0.882	0.368 to 0.412	0.326 to 0.369

[a]n: number of data points analyzed for each domain.

[b]MAPE: mean absolute percentage error.

Table 2. Validity of measuring heart rate with Polar OH1 and Fitbit Charge 3 devices. Data are aggregated to a single data point for each of the activity domains. Data are analyzed for all data points (column "Overall") and for each of the activity domains.

Device	Overall (n=111[a])	Rest (n=19[a])	Cycling light (n=19[a])	Cycling hard (n=19[a])	Treadmill (n=19[a])	Sprint cycling (n=18[a])	Sprint running (n=17[a])
Polar H10							
Heart rate (beats·min^{-1}), mean (SD)	106 (27)	62 (9)	89 (11)	119 (12)	124 (14)	128 (17)	120 (12)
Polar OH1							
Heart rate (beats·min^{-1}), mean (SD)	105 (27)	62 (9)	89 (10)	118 (12)	124 (14)	122 (17)	121 (15)
Mean bias (beats·min^{-1})	1	0	0	1	0	5	1
Limits of agreement (beats·min^{-1})	−8 to 10	0 to 1	−3 to 2	−2 to 4	−1 to 2	−8 to 18	−16 to 15
MAPE[b] (%)	0.6	0.2	0.7	−0.8	0.2	−4.1	0.5
Correlation coefficient	0.954	0.983	0.974	0.985	0.992	0.807	0.67
95% CI of correlation coefficient	0.953 to 0.955	0.982 to 0.984	0.971 to 0.977	0.083 to 0.987	0.992 to 0.993	0.795 to 0.819	0.651 to 0.687
Fitbit Charge 3							
Heart rate (beats·min^{-1}), mean (SD)	97 (26)	61 (9)	81 (10)	100 (21)	123 (10)	109 (23)	112 (11)
Mean bias (beats·min^{-1})	−9	−1	−7	−19	−2	−18	−8
Limits of agreement (beats·min^{-1})	−41 to 23	−2 to 0	−36 to 22	−63 to 26	−13 to 10	−63 to 26	−28 to 12
MAPE (%)	7.37	−1.5	−7	−15	−1	−13.9	−6
Correlation coefficient	0.888	0.884	−0.056	0.183	0.924	0.771	0.496
95% CI of correlation coefficient	0.885 to 0.890	0.876 to 0.891	−0.113 to 0.002	0.127 to 0.238	0.921 to 0.927	0.756 to 0.784	0.471 to 0.520

[a]n: number of data points analyzed for each domain.

[b]MAPE: mean absolute percentage error.

Validity of HR for Each Activity Type

The mean bias and LoA for Polar OH1 and the criterion device were consistent for visit 1 activities; however, the LoA was much wider during HIIE exercise (Figure 1). Similarly, Polar OH1 performed consistently across each segment of visit 1, with very low MAPE and a very good correlation (Table 1). The MAPE and correlation coefficients were worse during HIIE activities (Table 1; Figure 1). Fitbit Charge 3 performed well with a low MAPE (<5%) and had a good correlation during the rest and treadmill activities. Although the MAPE was low for sprint cycling and sprint running, the correlation was poor (Table 1). Validity was poor during light cycle exercise (r=0.272; MAPE=−7.1%) and was very poor during vigorous cycling (r=0.183; MAPE=−16.4%). The mean bias ranged from −21 to −1 beats·min^{-1} and, together with the LoA, are depicted in Figure 1. Validity across participants was consistent for OH1, where the MAPE was <5% for each participant (Table 3). The MAPE for each participant for Fitbit Charge 3 exceeded 5% in 7 of 19 participants (Table 3). Individual traces for each of the 3 devices during sprint cycling are shown in Figure 2.

Figure 1. Bland and Altman plots for unaveraged data across each activity domain. Subpart A shows data from Polar OH1, and subpart B shows data from Fitbit Charge 3. Solid blue line represents the mean bias, and blue dashed lines represent the limits of agreement.

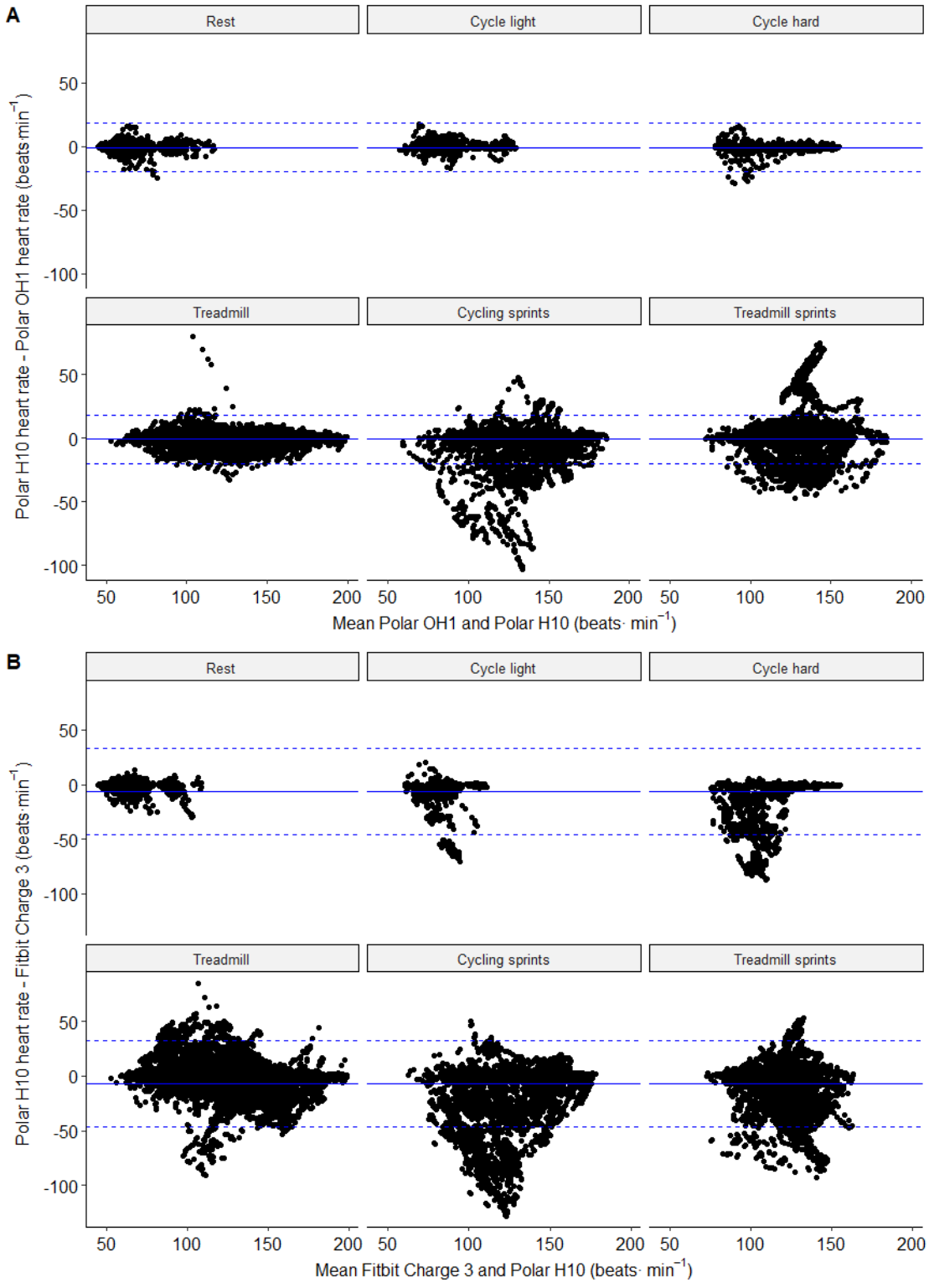
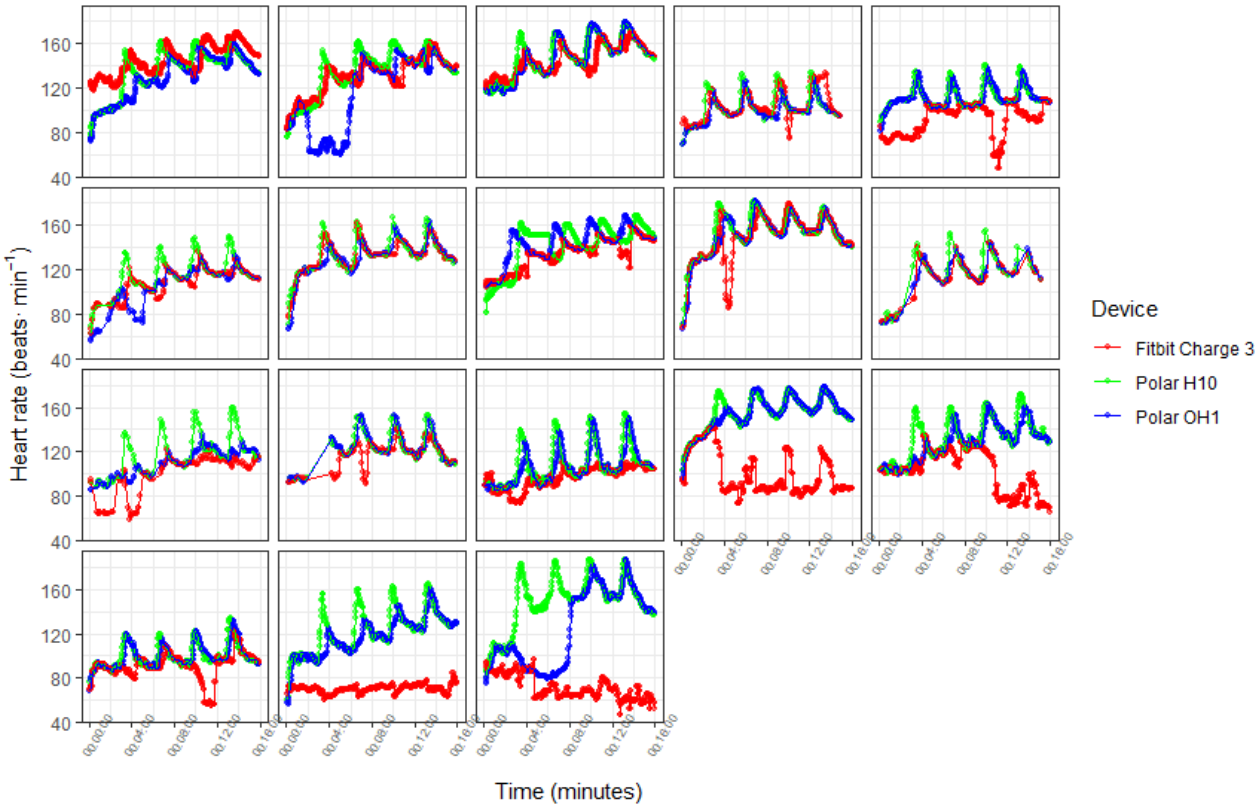


Table 3. The mean absolute percentage error for each participant for Polar OH1 and Fitbit Charge 3 across all available data points.

Participant	Polar OH1 (%)	Fitbit Charge 3 (%)
1	0.22	6.34[a]
2	−3.92	5.30[a]
3	1.37	2.80
4	0.06	2.06
5	0.31	4.01
6	1.70	1.53
7	0.22	2.35
8	−0.44	2.59
9	0.71	3.14
10	−0.09	2.83
11	0.59	-3.25
12	1.81	4.00
13	−0.10	3.91
14	−0.09	−5.43[a]
15	0.69	14.90[a]
16	0.87	5.97[a]
17	0.19	2.75
18	1.18	11.40[a]
19	2.58	15.80[a]
Overall	0.41	4.37

[a]Exceeds 5% mean absolute percentage error threshold.

Figure 2. Individual traces during sprint cycling. Green line represents Polar H10 (criterion device), the blue line represents Polar OH1, and the red line represents Fitbit Charge 3. Traces show 4 peaks in heart rate for each of the sprints followed by a recovery period of 3 minutes.



Discussion

Principal Findings

To our knowledge, this is the first study to assess the validity of Fitbit Charge 3 and Polar OH1 across a range of activity types, including HIIE or SIE. The main findings were that Polar OH1 showed good agreement in assessing HR versus the criterion measure (Polar H10) across activity domains or types in trial 1, whereas the validity of Fitbit Charge 3 was only acceptable during rest and treadmill activities. Fitbit Charge 3 performed particularly poorly during cycling exercise, where the mean bias ranged from −7 to −21 beats·min^{-1}, and LoA were very wide compared with other activity types. Finally, our data suggest that both Polar OH1 and Fitbit Charge 3 devices performed poorly during the visit 2 sprint cycling and sprint running protocols compared with the visit 1 activities.

The findings in this study suggest that Polar OH1 performs within acceptable tolerance limits for measuring HR during a range of activity types and intensities (ie, MAPE range 0%-4%). These findings are consistent with those of previous work [7]. Despite this, our data suggest that OH1 may be less capable of accurately measuring HR during sprint-based exercise.

Figure 2 shows the individual responses to sprint cycling for each of the 3 devices. Polar OH1 followed a similar trend to Polar H10 in 13 of the 18 available data sets. In contrast, Fitbit Charge 3 only followed a similar profile to Polar H10 in 5 of the 18 participant data sets. A typical observation across participants was a lower peak HR versus the criterion device and a slower response in HR change following the onset of sprint work. Given the popularity of HIIE and SIE training, these limitations may be crucial for assessing intensity accurately, and further development of the technology is required before it can be used accurately instead of ECG-type devices across all activity types. For everyday activities and for measuring HR in a range of activities, PPG-type devices may therefore be a more practical solution for estimating the intensity of activities. This may be particularly relevant in older, overweight or obese, or clinical populations, where wearing a chest strap may be off-putting or impractical.

This is the first published study to examine the validity of Fitbit Charge 3, although several studies have investigated the validity of its predecessors, Fitbit Charge HR (released 2015) and Fitbit Charge 2 (released 2016). We observed HR measurements that were within an acceptable percentage error range (0%-5%) during rest and the incremental treadmill test. However, Fitbit Charge 3 exhibited MAPE >5% during sprint running and during light, hard, and sprint cycling, with particularly large MAPE and mean bias observed during hard and sprint cycling. Previous studies have reported that Fitbit Charge HR and Fitbit Charge 2 underestimate HR in comparison with criterion devices during cycle-based activities [8,9,11,14]. In addition, increasing exercise intensity appears to increase Fitbit Charge HR and Charge 2 measurement errors [8-10,14,23-25]. Furthermore, Reddy et al [9] reported that Fitbit Charge 2 criterion-related validity was poor during cycling when transitioning swiftly from low to high intensity, which is in agreement with our cycling SIE findings, suggesting that Fitbit Charge model exhibits a measurement lag when HR increases rapidly. Therefore, under the conditions of this study, Fitbit Charge 3

performs as poorly as its predecessors when measuring HR during cycling activities and high-intensity exercise.

We can only speculate as to why Fitbit Charge 3 performed worse than Polar OH1 during cycling and high-intensity exercise in this study. Olstad and Zinner [26] previously suggested that wrist-based devices may be less sensitive to sudden changes in exercise intensity, such as when transitioning from low to high intensity, as experienced in SIE, because peripheral resistance is lower at the wrist, reducing changes in pulse pressure and disrupting blood pulse detection [16]. Therefore, the positioning of Polar OH1 may partially mitigate the poor signal detected at the wrist location. Furthermore, greater movement at the wrist may contribute to poor measurement accuracy [8,9,27], particularly during activities involving sustained hand and forearm muscle contractions [15], as may be experienced when gripping the handlebars during intense cycling. It is possible that Polar OH1 is less prone to movement artifacts because of a more robust strap design, which Spierer et al [15] previously identified as a potential contributor to differences in HR measurements among devices. In addition, Thomson et al [10] stated that hardware differences may affect the signal-to-noise ratio of the device. PurePulse technology of Fitbit Charge 3 may be more sensitive to interference than the technology of Polar OH1, which could have resulted in the less accurate HR measurements obtained in this study. Further advancements in technology are inevitable and will require further work from the scientific community to scrutinize devices and software advancements. Nevertheless, PPG-based devices provide an exciting opportunity to improve physical activity levels, promote adherence to exercise interventions, and drive behavior change.

This study included a variety of exercise intensities and both cycling and treadmill activities. In addition, few studies have examined the validity of consumer wearables in measuring HR during SIE, an increasingly prevalent exercise modality. However, despite providing novel insights into the accuracy of Fitbit Charge 3 and Polar OH1 in the detection of HR, this study has limitations. The sample size of this study, which is consistent with other similar studies [6,7,13,26,28-30], is small; therefore, the elements of the study are likely underpowered. This is particularly true for the HIIE type exercise, where a large variation in HR is observed, which is far from zero. As a result, it is inevitable that the associated LoAs are large and that a larger sample size is required to reduce the variation in the measurement. We refer readers to a new consensus statement by Mühlen et al [31] for further reading regarding sample size estimation for validation studies and for the design of validation studies in general. Furthermore, given the significant effort and time required to perform these types of studies, adequate funding is required from manufacturers to appropriately address the issue of sample size in validation studies. Other limitations include that the exercise tasks were conducted in a controlled laboratory environment, and the performance of these devices may differ when investigated under free-living conditions. This study, similar to many activity tracker validation studies, elected to recruit a cohort of individuals who were healthy and young to middle aged (25-56 years). Therefore, our findings cannot be generalized to clinical populations or different age groups. Furthermore, although not always the case [32], skin tone has been found to affect PPG signals [33] but was not accounted for in this study. Finally, our study design did not assess device acceptability, which should be considered if these devices were to be used in community-based interventions.

Conclusions

In conclusion, our data suggest that Polar OH1 is a suitable method for measuring HR during cycling, walking, and running activities within a healthy population. In contrast, data pertaining to Fitbit Charge 3 should be interpreted with caution, particularly during cycling activities. This may have significant implications for exercise training or rehabilitation purposes, where attainment of exercise intensity is a key aspect for cardiorespiratory fitness progression or where safety considerations exist. Furthermore, both PPG sensors evaluated in this study performed worse during the SIE activities. Given the rise in popularity of HIIE or SIE, we recommend that more traditional ECG/HR monitors are used when performing these activities. For the general population and scientific community to appropriately interpret PPG data, researchers should continue to assess the validity of new and existing devices among various populations and settings.

Acknowledgments

The authors would like to thank all participants who volunteered to participate in this study. DM, OG, DC and IM are supported by the European Union's INTERREG VA Programme, managed by the Special European Union Programmes Body. DC is also supported by a grant from Highlands and Islands Enterprise (HMS 9353763). KH is partly funded by a grant from Highlands and Islands Enterprise (HMS 9353763) and partly funded by Inverness and Highland City-Region Deal.

Authors' Contributions

DM, TG, IM, and DC conceptualized the study. DM, KH, AD, and DC collected the data. DM and DC analyzed the data and prepared the first draft of the paper. All authors have reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Wattbike resistance settings.
[DOCX File , 305 KB - mhealth_v9i3e25313_app1.docx]

Multimedia Appendix 2
Sprint running distance settings.
[DOCX File , 19 KB - mhealth_v9i3e25313_app2.docx]

Multimedia Appendix 3
Sprint running treadmill belt resistance (parachute size) settings.
[DOCX File , 19 KB - mhealth_v9i3e25313_app3.docx]

References

1. Brickwood KJ, 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) [FREE Full text] [doi: 10.2196/11819] [Medline: 30977740]
2. Aubert AE, Seps B, Beckers F. Heart rate variability in athletes. Sports Med 2003;33(12):889-919. [doi: 10.2165/00007256-200333120-00003] [Medline: 12974657]
3. Borresen J, Lambert MI. Changes in heart rate recovery in response to acute changes in training load. Eur J Appl Physiol 2007 Nov 9;101(4):503-511. [doi: 10.1007/s00421-007-0516-6] [Medline: 17687564]
4. Buchheit M. Monitoring training status with HR measures: do all roads lead to Rome? Front Physiol 2014;5:73 [FREE Full text] [doi: 10.3389/fphys.2014.00073] [Medline: 24578692]
5. Schneider C, Hanakam F, Wiewelhove T, Döweling A, Kellmann M, Meyer T, et al. Heart rate monitoring in team sports-a conceptual framework for contextualizing heart rate measures for training and recovery prescription. Front Physiol 2018 May 31;9:639 [FREE Full text] [doi: 10.3389/fphys.2018.00639] [Medline: 29904351]
6. Schubert MM, Clark A, De La Rosa AB. The polar OH1 optical heart rate sensor is valid during moderate-vigorous exercise. Sports Med Int Open 2018 Jun;2(3):E67-E70 [FREE Full text] [doi: 10.1055/a-0631-0920] [Medline: 30539120]
7. Hettiarachchi IT, Hanoun S, Nahavandi D, Nahavandi S. Validation of Polar OH1 optical heart rate sensor for moderate and high intensity physical activities. PLoS One 2019 May 23;14(5) [FREE Full text] [doi: 10.1371/journal.pone.0217288] [Medline: 31120968]
8. 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 Mar;50(3):624-633. [doi: 10.1249/MSS.0000000000001471] [Medline: 29189666]
9. 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 Dec 10;6(12) [FREE Full text] [doi: 10.2196/10338] [Medline: 30530451]
10. 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 Jun 18;37(12):1411-1419. [doi: 10.1080/02640414.2018.1560644] [Medline: 30657025]
11. 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 Feb 28;13(2) [FREE Full text] [doi: 10.1371/journal.pone.0192691] [Medline: 29489850]
12. Weston KS, Wisløff U, Coombes JS. High-intensity interval training in patients with lifestyle-induced cardiometabolic disease: a systematic review and meta-analysis. Br J Sports Med 2014 Aug 21;48(16):1227-1234. [doi: 10.1136/bjsports-2013-092576] [Medline: 24144531]
13. Khushhal A, Nichols S, Evans W, Gleadall-Siddall D, Page R, O'Doherty A, et al. Validity and reliability of the apple watch for measuring heart rate during exercise. Sports Med Int Open 2017 Oct 18;1(6):206-211 [FREE Full text] [doi: 10.1055/s-0043-120195] [Medline: 30539109]
14. Jo E, Lewis K, Directo D, Kim MJ, Dolezal BA. Validation of biofeedback wearables for photoplethysmographic heart rate tracking. J Sports Sci Med 2016 Sep;15(3):540-547 [FREE Full text] [Medline: 27803634]
15. Spierer DK, Rosen Z, Litman LL, Fujii K. Validation of photoplethysmography as a method to detect heart rate during rest and exercise. J Med Eng Technol 2015 Jun 26;39(5):264-271. [doi: 10.3109/03091902.2015.1047536] [Medline: 26112379]
16. Johnson JM. Physical training and the control of skin blood flow. Med Sci Sports Exerc 1998 Mar;30(3):382-386. [doi: 10.1097/00005768-199803000-00007] [Medline: 9526883]
17. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
18. Gilgen-Ammann R, Schweizer T, Wyss T. RR interval signal quality of a heart rate monitor and an ECG Holter at rest and during exercise. Eur J Appl Physiol 2019 Jul 19;119(7):1525-1532. [doi: 10.1007/s00421-019-04142-5] [Medline: 31004219]
19. PulseWatch. URL: https://iccir919.github.io/pulseWatch/public/index.html [accessed 2021-03-05]

20. Gellish RL, Goslin B, Olson R, McDonald A, Russi G, Moudgil V. Longitudinal modeling of the relationship between age and maximal heart rate. Med Sci Sports Exerc 2007 May;39(5):822-829. [doi: 10.1097/mss.0b013e31803349c6] [Medline: 17468581]
21. Bland MJ, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet 1986 Feb;327(8476):307-310. [doi: 10.1016/s0140-6736(86)90837-8]
22. Shcherbina A, Mattsson C, Waggott D, Salisbury H, Christle J, 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 May 24;7(2):3 [FREE Full text] [doi: 10.3390/jpm7020003] [Medline: 28538708]
23. Wang R, Blackburn G, Desai M, Phelan D, Gillinov L, Houghtaling P, et al. Accuracy of wrist-worn heart rate monitors. JAMA Cardiol 2017 Jan 01;2(1):104-106. [doi: 10.1001/jamacardio.2016.3340] [Medline: 27732703]
24. 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 Mar 16;5(3):e34 [FREE Full text] [doi: 10.2196/mhealth.7043] [Medline: 28302596]
25. Dondzila CJ, Lewis CA, Lopez JR, Parker TM. Congruent accuracy of wrist-worn activity trackers during controlled and free-living conditions. Int J Exerc Sci. 2018. URL: https://digitalcommons.wku.edu/cgi/viewcontent.cgi?article=2251&context=ijes [accessed 2021-03-05]
26. Olstad BH, Zinner C. Validation of the Polar OH1 and M600 optical heart rate sensors during front crawl swim training. PLoS One 2020;15(4) [FREE Full text] [doi: 10.1371/journal.pone.0231522] [Medline: 32298322]
27. Horton JF, Stergiou P, Fung T, Katz L. Comparison of Polar M600 optical heart rate and ECG heart rate during exercise. Med Sci Sport Exerc 2017;49(12). [doi: 10.1249/mss.0000000000001388]
28. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR. Accuracy of PurePulse photoplethysmography technology of Fitbit Charge 2 for assessment of heart rate during sleep. Chronobiol Int 2019 Jul 16;36(7):927-933. [doi: 10.1080/07420528.2019.1596947] [Medline: 30990098]
29. Tedesco S, Sica M, Garbay T, Barton J, OrFlynn B. A comprehensive comparison of commercial wrist- worn trackers in a young cohort in a lab- environment. In: Proceedings of the IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). 2018 Presented at: IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops); March 19-23, 2018; Athens, Greece. [doi: 10.1109/PERCOMW.2018.8480386]
30. Wallen MP, Gomersall SR, Keating SE, Wisløff U, Coombes JS. Accuracy of heart rate watches: implications for weight management. PLoS One 2016 May 27;11(5) [FREE Full text] [doi: 10.1371/journal.pone.0154420] [Medline: 27232714]
31. Mühlen JM, Stang J, Skovgaard LE, Judice PB, Molina-Garcia P, Johnston W, et al. Recommendations for determining the validity of consumer wearable heart rate devices: expert statement and checklist of the INTERLIVE Network. Br J Sports Med 2021 Jan 04 [FREE Full text] [doi: 10.1136/bjsports-2020-103148] [Medline: 33397674]
32. Sañudo B, De Hoyo M, Muñoz-López A, Perry J, Abt G. Pilot study assessing the influence of skin type on the heart rate measurements obtained by photoplethysmography with the apple watch. J Med Syst 2019 May 22;43(7):195. [doi: 10.1007/s10916-019-1325-2] [Medline: 31119387]
33. Butler MJ, Crowe JA, Hayes-Gill BR, Rodmell PI. Motion limitations of non-contact photoplethysmography due to the optical and topological properties of skin. Physiol Meas 2016 Apr 21;37(5):N27-N37. [doi: 10.1088/0967-3334/37/5/n27]

Abbreviations

ANS: autonomic nervous system
ECG: electrocardiogram
HIIE: high-intensity interval exercise
HR: heart rate
IPAQ-SF: International Physical Activity Questionnaire-Short Form
LoA: limits of agreement
MAPE: mean absolute percentage error
PPG: photoplethysmography
SIE: sprint interval exercise

Edited by L Buis; submitted 27.10.20; peer-reviewed by G Abt, A Samuel; comments to author 06.12.20; revised version received 08.01.21; accepted 18.02.21; published 25.03.21.

Please cite as:
Muggeridge DJ, Hickson K, Davies AV, Giggins OM, Megson IL, Gorely T, Crabtree DR
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
URL: https://mhealth.jmir.org/2021/3/e25313
doi:10.2196/25313
PMID:33764310

Viewpoint

Using Fitbit as an mHealth Intervention Tool to Promote Physical Activity: Potential Challenges and Solutions

Guilherme M Balbim[1*], PhD; Isabela G Marques[2], PhD; David X Marquez[1], PhD; Darshilmukesh Patel[3], BSc; Lisa K Sharp[4], PhD; Spyros Kitsiou[5*], PhD; Sharmilee M Nyenhuis[6*], MD

[1]Department of Kinesiology and Nutrition, College of Applied Health Sciences, University of Illinois at Chicago, Chicago, IL, United States

[2]Rheumatology Division, School of Medicine, University of Sao Paulo, Sao Paulo, Brazil

[3]Department of Medicine, College of Medicine, University of Illinois at Chicago, Chicago, IL, United States

[4]College of Pharmacy, University of Illinois at Chicago, Chicago, IL, United States

[5]Department of Biomedical and Health Information Sciences, College of Applied Health Sciences, University of Illinois at Chicago, Chicago, IL, United States

[6]Division of Pulmonary, Critical Care, Sleep and Allergy, Department of Medicine, University of Illinois at Chicago, Chicago, IL, United States

[*]these authors contributed equally

Corresponding Author:
Spyros Kitsiou, PhD
Department of Biomedical and Health Information Sciences
College of Applied Health Sciences
University of Illinois at Chicago
1919 W Taylor St (MC 530)
Chicago, IL, 60612
United States
Phone: 1 312 355 3519
Email: skitsiou@uic.edu

Abstract

Consumer-based physical activity (PA) trackers, also known as wearables, are increasingly being used in research studies as intervention or measurement tools. One of the most popular and widely used brands of PA trackers is Fitbit. Since the release of the first Fitbit in 2009, hundreds of experimental studies have used Fitbit devices to facilitate PA self-monitoring and behavior change via goal setting and feedback tools. Fitbit's ability to capture large volumes of PA and physiological data in real time creates enormous opportunities for researchers. At the same time, however, it introduces a number of challenges (eg, technological, operational, logistical), most of which are not sufficiently described in study publications. Currently, there are no technical reports, guidelines, nor other types of publications discussing some of these challenges and offering guidance to researchers on how to best incorporate Fitbit devices in their study design and intervention to achieve their research goals. As a result, researchers are often left alone to discover and address some of these issues during the study through "trial and error." This paper aims to address this gap. Drawing on our cumulative experience of conducting multiple studies with various Fitbit PA trackers over the years, we present and discuss various key challenges associated with the use of Fitbit PA trackers in research studies. Difficulties with the use of Fitbit PA trackers are encountered throughout the entire research process. Challenges and solutions are categorized in 4 main categories: study preparation, intervention delivery, data collection and analysis, and study closeout. Subsequently, we describe a number of empirically tested strategies used in 4 of our interventional studies involving participants from a broad range of demographic characteristics, racial/ethnic backgrounds, and literacy levels. Researchers should be prepared to address challenges and issues in a timely fashion to ensure that the Fitbit effectively assists participants and researchers in achieving research and outcome goals.

(JMIR Mhealth Uhealth 2021;9(3):e25289) doi:10.2196/25289

KEYWORDS

physical activity; fitness trackers; Fitbit; smartphones; interventional studies; adults; older adults; wearable; intervention

Background

The 2018 Physical Activity Guidelines for Americans Scientific Advisory Committee report shows strong evidence of the role of physical activity (PA) in reducing the risk of chronic diseases [1]. PA guidelines recommend that adults should engage in 150 to 300 minutes of moderate-intensity, or 75 to 150 minutes of vigorous-intensity, aerobic PA each week. However, only 53.3% of adults in the United States meet this guideline [1,2], while globally, 27.5% of adults are insufficiently active [3].

In 2019, fitness trackers and smartwatches (eg, Fitbit, Apple watch, Garmin), hereby termed PA trackers, were the number one fitness trend, with 19% of Americans owning at least one [4]. PA trackers measure a variety of variables including daily steps, intensity-specific minutes of PA, heart rate, calorie expenditure, stairs climbed, sedentary behavior, sleep duration, and sleep performance [5,6]. These devices are worn at various body locations, with hip, wrist, and thigh being the most common placements [6]. These commercially available PA trackers have been used in research studies to promote and monitor PA and in clinical practice for patients with chronic illness(es) where PA has shown to have a positive impact [7].

Among several brands of commercial PA trackers, Fitbit has emerged as one of the most popular in the wearable industry, having sold more than 100 million devices worldwide [8]. According to a recent International Data Corporation report on wearable devices, Fitbit is one of the leading companies in the wearable PA tracker space and has a large user base with more than 28 million active users [9]. A similar dominant trend is also observed in research, where Fitbits are the most frequently used devices, particularly in interventional studies that focus on promoting PA and other healthy lifestyle behaviors [10]. A search on the Fitabase library [11], which maintains a list of published research studies that have used Fitbit PA trackers, yielded over 200 validity studies and 172 interventional studies (this list is not exhaustive). Despite the large number of studies that have used Fitbit devices for research purposes, no prior publication has offered detailed insights regarding some of the challenges and limitations associated with their use in research studies for data collection, behavior change, and outcome assessment. Furthermore, there are no published technical reports or best practice guidelines to support researchers on how to best incorporate these devices in their study design and intervention. As a result, researchers are often left alone to discover and address some of these issues through “trial and error.” This paper aims to address this important gap by describing key challenges and solutions associated with the use of Fitbit PA trackers. While we recognize that this paper might be applicable to other commercially available PA trackers, a direct comparison with them or attempt to explain how these challenges and solutions may apply to other PA trackers goes beyond the scope of this paper.

Over the years, our research team has led several research studies utilizing Fitbits as intervention tools. Based on our collective experience with these studies, we identified several challenges that researchers need to take into account when designing and conducting studies that utilize Fitbit PA trackers. These challenges arise in distinct stages of a study, including study preparation, intervention delivery, and data management, and may have a significant impact on the conduct of the study and analysis of results. With the experience acquired, we describe empirically tested strategies and solutions that were effective in our interventions that included a broad range of demographic characteristics, racial/ethnic backgrounds, and educational levels.

Description of the Studies

Between our team, we have conducted 4 separate studies utilizing Fitbit PA trackers. In this paper, we reference our experience from these studies to describe the key challenges associated with the use of Fitbits in research and subsequently summarize the strategies and remedies we used in each of these studies to address each challenge. Briefly, the included studies are (1) BAILA TECH [12,13], a Latin dance program for middle-aged and older Latinos; (2) ACTION [14], a lifestyle PA program for African American women with asthma; (3) Virtual Coach Study [15], a text-based, goal-setting intervention for adults with chronic diseases; and (4) iCardia4HF [16], a patient-centered, mobile health (mHealth) technology intervention to promote self-care in adult patients with chronic heart failure (HF). Table 1 depicts the main characteristics of each study.

Table 1. Characteristics of the studies.

Characteristics	Study			
	BAILA TECH [12,13]	ACTION [14]	Virtual Coach [15]	iCardia4HF [16]
Study Design	Single-group pre-post feasibility	Randomized controlled trial	Single-group pre-post exploratory	Randomized controlled trial
Participants				
Sample size	20	53 ($N_{intervention}$=25, $N_{control}$=28)	30	25 ($N_{intervention}$=11, $N_{control}$=14)
Age (years), mean (SD)	67.0 (7.1)	43.4 (12.2)	47.0 (9.5)	56.0 (8.3)
Gender (female), n (%)	15 (75)	53 (100)	21 (75)	11 (44)
Race/ethnicity	Latinos	African American	African American (79%) and Latinos (18%)	African American (92%)
PA[a] status	Low-active	Low-active	Sedentary	Sedentary
Health status	Healthy	Uncontrolled asthma (Asthma Control Test <20)	Chronic diseases (eg, hypertension, asthma, type 2 diabetes)	Chronic heart failure
Intervention				
Type	BAILAMOS dance program [17]	Modified version of the Women's Lifestyle Physical Activity Program [18]	Remote goal setting and self-monitoring	Patient-centered mHealth[b] technology intervention to promote adherence to self-care in patients with chronic heart failure
mHealth components	Fitbit Charge 2, Fitbit mobile app, text messages	Fitbit Charge HR, Fitbit mobile app, text message	Fitbit tracker (participants were given a choice of the following trackers: Alta, Charge 2, or Charge 3), Fitbit mobile app, text messages	Fitbit Charge 2, Fitbit mobile app, Withings Body Cardio scale, Withings Blood Pressure cuff, Health Mate mobile app, Heart Failure Health Storylines mobile app, text messages
Length (weeks)	16	24	8	8
Frequency	2x/week	5 monthly group sessions (barriers for exercising with asthma, Fitbit usage, and review	N/A[c]	Daily
Duration	30 min Fitbit instructional session + 60 min dance + 30 min technological troubleshooting (optional)	120 min	N/A	N/A
Comparator	N/A	Fitbit Charge HR + PA educational material	N/A	Usual care (outpatient follow-up at the Heart Failure Center 7 days after hospital discharge and every 3 months thereafter)
Outcomes	Feasibility and acceptability of the BAILA TECH intervention, PA levels (eg, steps/day, engagement in moderate-to-vigorous PA/week)	Feasibility and acceptability of the modified Women's Lifestyle Physical Activity Program, PA levels (eg, steps/day, low PA, moderate-to-vigorous PA/week, sedentary time), Asthma Control questionnaire [19], Mini-Asthma Quality of Life [19]	Feasibility and acceptability of the Virtual Coach intervention, average steps per week	Feasibility and acceptability of the intervention, heart failure self-care [20], PA levels (steps/day and moderate-to-vigorous PA/week)

[a]PA: physical activity.

[b]mHealth: mobile health.

[c]N/A: not applicable.

The BAILA TECH study [12,13] delivered the BAILAMOS Latin dance program tailored to middle-aged and older Spanish-speaking Latinos (for details, see Marquez et al [17]) infused with mHealth components [12,13]. The BAILAMOS dance program had 60-minute classes twice a week for 16 weeks and consisted of 4 Latin dance styles (Merengue, Bachata, Cha Cha Cha, and Salsa). For BAILA TECH [12,13], before each dance class, participants also attended 30-minute Fitbit instructional sessions. After each dance class, the research team was available for 30 minutes in case participants needed more assistance for troubleshooting technological issues. Participants were asked to wear their Fitbit for 2 weeks prior to intervention start, throughout the intervention period (16 weeks), and 1 week after the intervention, totaling 19 weeks.

The ACTION study delivered a modified version of the Women's Lifestyle Physical Activity Program [14]. The original program was a 24-week, evidence-based, moderate-intensity walking intervention designed for urban mid-life African American women including (1) tailored walking prescription, (2) PA self-monitoring tool (pedometer), (3) group discussions to address goals and barriers, and (4) motivational telephone calls. The program was adapted to account for the asthma barriers to PA. The modified version was a 24-week intervention and included (1) a Fitbit Charge HR device, (2) one 2-hour educational session at the start of the intervention dedicating approximately 45 minutes to asthma education (based on the American Lung Associations Asthma Basics course) and approximately 30 minutes to Fitbit orientation, (3) weekly motivational text messages (average of 3 per week), and (4) monthly 2-hour group sessions to address barriers of exercising with asthma, answer Fitbit-related questions, review personalized step goals, and walk as a group. Participants were asked to wear their Fitbit throughout the intervention period (24 weeks).

The Virtual Coach study [15] delivered remote goal setting based on text messages and self-monitoring. Each participant was assigned to a coach and set individual daily goals with their coaches. Weekly goals outlined specific days, times, and number of steps for each of the 8 weeks of intervention. Participants were asked to evaluate how confident they were on reaching each goal (scale of 0-10). If a participant rated the confidence less than 8, coaches texted with the participant to modify the goal until confidence ratings were between 8 and 10 (highest confidence). Participants were asked to wear their Fitbit throughout the intervention period (8 weeks).

The iCardia4HF study [16] delivered a patient-centered, mHealth technology intervention to promote adherence to self-care in patients with chronic HF. The intervention lasted for 8 weeks and consisted of (1) the Heart Failure Health Storylines mobile app, (2) 3 connected health devices (Fitbit Charge 2, Withings Body Cardio scale, and blood pressure monitor) that connect to the app to facilitate self-monitoring of multiple HF-related parameters (eg, heart rate, weight, blood pressure, and PA), and (3) individually tailored text messages targeting health beliefs, self-efficacy, and HF knowledge. Participants were asked to wear the Fitbit for 8 weeks for 10 or more hours per day to monitor and gradually increase their daily steps.

Challenges for Fitbit Use in Interventions and Potential Strategies and Solutions

Overview

Challenges and solutions are classified into 4 main categories: study preparation, intervention delivery, data management, and study closeout. Two of these categories are subdivided according to the type of challenge presented. Within study preparation, we identified difficulties related to the Fitbit app and the user. During the intervention delivery, challenges were related to the interaction of participants with the Fitbit app and Fitbit tracker. We adopted this structure with the intention to provide a logical flow parallel to the conduct of a study (ie, study design, execution or management, and study closeout). Researchers should be aware that challenges related to intervention delivery and data management will occur concurrently during the study.

Study Preparation

Selecting a PA tracker that aligns with study goals, participant preferences, and the data collection environment is essential and one of the first challenges a researcher might face early in the research process. Similar to other vendors, Fitbit has a large variety of trackers with different monitoring capabilities (eg, heart rate, sleep, automatic exercise tracking, geolocation), features (eg, automated notifications of sedentary behavior), and aesthetics (eg, large vs smaller screen and bands). All Fitbit models require that participants own a smartphone or other Bluetooth-enabled devices (eg, tablet) to synchronize data from the Fitbit tracker to the Fitbit app. Researchers should review the characteristics of available Fitbit models and, if possible, conduct formative research to gather data that align participants' preferences with the study's needs. For example, if exercise sessions or quality and duration of sleep are of direct interest to a study, then clip-on models (Fitbit Zip) that cannot automatically track these behaviors will not be a good fit for the purposes of the study. Successful strategies we adopted in our studies included focus groups [21], field usability testing followed by a pre-pilot study to assess the feasibility of using Fitbit trackers as part of the study intervention [22], and surveys [23] to evaluate participants' preferences and perceptions about the use of Fitbit for PA self-monitoring. For example, before the main trial in 2 studies (ie, ACTION [14] and BAILA TECH [12,13]), we obtained input from focus groups regarding their familiarity with using PA trackers and if they would feel comfortable using these devices. In another study (ie, Virtual Coach Study [15]), we allowed the first few participants to try on different Fitbit models (Alta, Charge 2, and Charge 3) and solicited input from study participants to decide which model should be used in the preliminary study. In the iCardia4HF study [23], we conducted a cross-sectional survey to assess the perceptions and attitudes of patients with HF towards using various Fitbit trackers and mobile apps for self-monitoring of PA and other important measures (eg, heart rate).

After selection of the proper Fitbit model, researchers must decide how the Fitbit data will be collected from study participants during the study. Fitbit has a website dashboard in place that allows users to view and export cumulative data pertaining to their daily steps, intensity of activity, sedentary

minutes, and other device-based measures. However, this dashboard is not suitable for research studies, particularly those with large sample sizes, because it does not allow concurrent access and exportation of data from multiple study participants. Data exportation is limited to the historic data of 1 participant at a time. Also, the Fitbit dashboard does not allow research staff to download intraday data such as minute-by-minute heart rate or steps, which can be used in a study as proxy measures to calculate Fitbit wear time (for more details, see the Data Management section) or to tailor the content of the intervention. For small studies (n≤20), manually checking and exporting data for 1 participant at a time may not be an important issue. However, for larger studies, this poses a significant barrier. Large studies require adequate strategies and robust mechanisms to remotely and automatically gather large amounts of incoming data from all participants in a single platform. Research teams have the option to utilize third-party commercial data collection platforms (eg, Fitabase) or research platforms (eg, iCardia [24], Mytapp) in order to remotely collect Fitbit data from participants' accounts using Fitbit's Web application programming interface (API). These platforms provide secure data acquisition and management tools that facilitate remote data collection from Fitbit PA trackers without the need for study participants to return the devices to researchers for data extraction. Another potential solution (if these platforms cannot support the needs of the study) is for researchers to develop their own data collection interfaces with the Fitbit cloud servers and data collection management environment for the purpose of their study using Fitbit's API. However, this approach is time-consuming and requires sufficient programing expertise, including testing before utilization in the main study.

In our experience, the use of research platforms (eg, iCardia) that remotely collect Fitbit data from participants' accounts proved to be an effective, reliable, and cost-effective method. The iCardia platform was successfully implemented in 3 of the 4 studies [9,10,25], allowing our research team to view and export Fitbit data from multiple participants and also to deliver personalized text messages based on the incoming data (for more details about the platform, please see Kitsiou et al [24]). iCardia is a secure, password-protected system that is hosted in a Health Insurance Portability and Accountability Act (HIPAA)-compliant server at the University of Illinois at Chicago. Currently, it is used in several studies funded by the National Institutes of Health [25-27]. iCardia provides researchers with a number of visualization and analytics tools (eg, graphs, charts, and dashboard reports) to facilitate monitoring of Fitbit data in real time (eg, steps, sleep, heart rate), thus alleviating the need for the tedious and costly return of devices to researchers.

Another research-based platform adopted by one of our studies (ie, Virtual Coach [15]) is the Mytapp texting platform. Mytapp is a low-cost, multilingual, adaptive text messaging application developed at the University of Illinois at Chicago for health behavior interventions. It includes numerous functions including individual and group messaging, scheduled messaging, recurrent messaging, surveys (with responses based on simple logic statements), message personalization, and data collection pertaining to number of daily steps (only) from Fitbit devices. Mytapp is currently utilized by numerous National Institutes of Health–funded studies and was utilized by the Virtual Coach study [15].

Fitabase is another well-known cloud-based platform that extracts and aggregates participants' Fitbit data [28] and has been utilized in numerous intervention studies. Although Fitabase cannot send text messages to study participants via SMS, it has the ability to send notifications to Fitbit smartwatch devices (eg, Versa 2 and 3). None of our studies implemented Fitabase as a data collection environment. Nonetheless, a search with the descriptors "intervention AND Fitabase" on PubMed, Scopus, and ScienceDirect indicated 24 intervention studies using Fitabase since 2018.

Another challenge in the preparation phase is ensuring participants' anonymity and privacy when creating Fitbit accounts. Fitbit accounts require a unique email address for each user. While collecting unprecedented amounts of data and conducting interventions with mHealth tools, researchers should be concerned with protecting the privacy of research participants [29]. Additional concerns arise because some mHealth technologies (eg, mobile apps) may transfer participants' data to third-party companies for marketing purposes [29]. Therefore, if using third-party platforms, researchers are advised to check whether the platform fulfills all the necessary elements to ensure participants' anonymity, privacy, and HIPAA compliance.

Our experiences directed us to a 2-step solution: (1) setting up anonymous commercial or institutional email accounts and (2) setting up deidentified Fitbit accounts for the purpose of the study. This individualized email address will then be linked to a Fitbit account in which personal identifiers should not be used. The same procedure should be adopted if using an authorized third-party data collection environment. Importantly, research staff and participants need to be advised that the email and Fitbit accounts should be used strictly for the study's purpose. We suggest limiting the number of research staff responsible for linking personal identifiers with email and Fitbit account's user information. Moreover, account information should be stored in protected computers within encrypted documents. This strategy was successfully implemented within our 4 studies.

Another critical step before intervention delivery is calibrating the Fitbit according to the participants' characteristics (ie, height, weight, and stride length) and study needs (eg, exercise bout length, PA goals). This process should be conducted on an individual basis with the information inserted in the participant's Fitbit account. Research staff who have access to participants' personal identifiers can process the calibration at Fitbit's website utilizing participants' credentials (ie, anonymous Fitbit accounts). Importantly, participants should receive the Fitbit only after the calibration process is concluded. Specifically, ACTION's study [14] participants had their stride length measured at baseline data collection. A research staff member entered the stride length information for each participant before distributing the Fitbits. If it is not possible to measure participants' stride length, information on weight and height should be entered for each participant since this information acts as a proxy for stride length. In 1 of our studies (ie, BAILA

TECH [12,13]), participants provided self-reported weight and height at baseline to calibrate their Fitbits.

Table 2 displays the challenges and solutions or strategies at the study preparation stage.

Table 2. Challenges and solutions or strategies for utilizing the Fitbit physical activity (PA) tracker in health interventions at the study preparation stage.

Challenges	Strategies and solutions
Tracker and app related	
Selecting a PA tracker model that aligns with study goals and participant preferences	Select the proper model based on study goals and outcomes, using Fitbit's online comparison tool [30]; conduct formative research (if possible) with potential participants to assess ease of use, comfort, and preferences.
Selecting and setting up the data collection environment for the Fitbit data	Choose one of the following options: third-party or other existing research platforms (eg, Fitabase, iCardia), developing your own data collection platform for the purpose of your study, or using Fitbit's platform and dashboard to export individual data.
Setting up Fitbit study procedures	Develop a Fitbit manual of operations for research staff (setting up, pairing, and calibrating devices); devote sufficient time for training and also plan for frequent retraining of research staff on how to perform various Fitbit tasks and troubleshoot common issues (eg, Fitbit mobile app installation and setup, device pairing, and syncing), because of the numerous changes and updates that may occur during the study in the Fitbit app and devices; ensure anonymity of study participants by creating anonymous institutional or commercial email and Fitbit accounts.
Setting up and pairing participants Fitbit: poor internet connection; participants' smartphone and Fitbit compatibility issues (eg, outdated operating system, insufficient memory space, firmware updates)	Conduct the Fitbit set-up process with the assistance of trained research staff, test the quality of the internet connection at the location where the Fitbits will be set up, and check participants' phone compatibility with Fitbit at enrollment to determine if updates are needed or if participant needs a study (loaner) phone.
Fitbit calibration	Collect data on participants' weight, height, and if possible, stride length and add the information to the Fitbit app before giving the Fitbit to the participant; personalize PA goals according to the study's needs (eg, exercise bout length, goals) when possible.
Participant related	
Participants' unfamiliarity with Fitbit technology (unfamiliarity with PA tracker and app, concerns about privacy)	Develop a user manual according to participants' reading level explaining how to use the device and mobile app in the context of the study; conduct Fitbit orientation sessions with study participants; consider a run-in period (eg, 1 week) to allow participants to familiarize themselves with the technology; identify a superuser (eg, family member or research assistant) to assist with and troubleshoot technology issues; assure participants that procedures to ensure anonymity and privacy are in place.

Intervention Delivery

Before participants receive the Fitbits, the research staff need to be familiar with the Fitbit tracker and mobile app. To this end, staff members have to undergo sufficient training and experiential learning sessions on how Fitbit works and how to troubleshoot common issues (eg, Fitbit mobile app installation and setup, device pairing, and syncing). Given the frequent changes that may occur in the Fitbit app and devices during the study, researchers should also plan to have frequent retraining sessions. In addition to developing the training material, a manual of operations should be created and made available to all research staff that will interact with research participants. This strategy was successfully implemented with research staff in all 4 studies [12-16]. Issues with either the tracker or the mobile app will undoubtedly occur. Solving these issues as soon as possible is crucial to prevent data loss, which could potentially harm outcome assessment, fidelity, or feedback strategies that are dependent upon Fitbit data. Maintaining log sheets to record the date and characteristics of any technical issues occurring during the study and remedies applied to address them is strongly recommended.

Although technology use is familiar to many study participants, some segments of the population are not familiar with technological devices such as Fitbits and associated mobile apps. It is essential to be prepared for issues that may arise in relation to participants' interaction with the Fitbit tracker and mobile app. Before launching the intervention, it is highly advisable to conduct an orientation session to teach participants how to use the Fitbit tracker and mobile app in the context of the study. Our experiences showed that conducting orientation sessions is essential for the success of the intervention. We carried out orientation sessions with middle-aged and older Latinos (ie, BAILA TECH [12,13]), African American women with asthma (ie, ACTION [14]), older adults with HF (ie, iCardia4HF [16]), and low-income, minority adults with chronic diseases (ie, Virtual Coach [15]). Importantly, we also noticed that educational classes during the intervention were essential to enhance participants' engagement and satisfaction with Fitbit [31].

Additionally, the research team should distribute a user manual containing a guide on how to execute basic tasks required by the study, troubleshooting of common issues, and answers to frequently asked questions (Multimedia Appendix 1). The Fitbit User Manual available online [32] is a great resource to guide the development of a customized manual. Our studies utilized the Fitbit User Manual as a template. In BAILA TECH [12,13],

ACTION [14], and iCardia 4HF [16], we added images and more in-depth information of Fitbit features that were essential to achieve the intervention goals and for the success of data collection (eg, syncing, charging). It is essential to use plain and accessible language according to participants' educational attainment and literacy levels to facilitate comprehension of the manual. For example, in the BAILA TECH study [12,13], the manual was developed at a sixth-grade reading level.

Moreover, we recommend a run-in period (eg, 1 week) prior to baseline PA data collection to allow participants to familiarize themselves with the Fitbit device and its associated mobile app before the intervention starts. Run-in periods are useful in interventions with the use of technology since individuals often adopt and discontinue technology use at a different pace [33] and can flag the need for adaptations. If the study plans to collect PA data, the run-in period is essential. The first week is then disregarded, and the second week can be used for baseline PA assessment. This strategy was successfully adopted in the BAILA TECH study [12,13].

Identification of a superuser (eg, family member or research assistant) to assist with daily tasks and troubleshoot issues is also suggested. For example, the ACTION [14] and iCardia4HF [16] studies had 2 research assistants available to assist participants via text messages or phone, and if needed, in-person meetings were scheduled to troubleshoot. Nevertheless, researchers should be aware that providing assistance on an individual basis will likely increase participant-research staff interaction, which should be acknowledged as a potential confounder. In the ACTION [14] study, participants in the control group also received Fitbits; in such cases, the researcher should ensure that both groups receive equal technical support (ie, availability of research staff to troubleshoot) to avoid imbalances between the 2 groups.

The next step is intervention delivery, which presents several challenges. We classified these challenges in 2 categories: (1) participant-mobile app interaction and (2) participant-Fitbit tracker interaction. One of the most common issues we experienced in all 4 studies with participant-mobile app interaction was related to infrequent syncing of the Fitbit tracker with the Fitbit mobile app, which can lead to data loss if not identified and addressed in a timely manner. Fitbit trackers can hold up to 7 days of intraday data and up to 31 days of cumulative data in their memory without syncing. When a user syncs, the data are transferred through the phone to the cloud-based server, and the tracker's memory is flushed to create space for the new data. If the user does not sync within 7 days, the intraday data of the first day since the last sync are erased to create space for the new data of the eighth day, and so on. Thus, the process of syncing becomes critically important in interventions using Fitbits, especially if Fitbit is used as both an intervention and outcome assessment tool.

Without syncing, the research team does not have access to participants' data and runs the risk of data loss. Detecting participants who have not synced periodically (maximum of 7 days) and addressing the issue as soon as possible are very important and quite challenging in larger studies. In general, participants either forget to sync or forget how to go through the syncing process. The Fitbit app offers a "sync automatically" option as long as the Bluetooth and internet connections are enabled and the app is active or running in the background. However, we noticed that many participants tend to turn off their Bluetooth or the app to save battery on their phone and then forget to turn them back on or access the app to sync. To address this challenge, study participants should have easy access to information on how to sync the tracker with the mobile app (eg, manual). We also recommend that the research staff closely monitor participants' data daily (eg, via the platform). It is easy to detect the participants who have not synced recently or if their Fitbit battery is low and subsequently send a notification (eg, via a text message, phone call, or email) to sync or charge their battery. Fitbit Charge 2 and other newer models display a battery notification on the tracker's screen when the battery energy is below 20%. Therefore, a reminder from the research team may not be needed. In the BAILA TECH [12,13] and ACTION [14] studies, research assistants were tracking participants' synced Fitbit data daily. After 3 days without syncing, research assistants contacted the participant via SMS reminding them to sync their Fitbit. The BAILA TECH study [12,13] delivered an intervention that allowed research assistants to meet participants twice a week, which made it easier to remind participants to sync their Fitbits in person. Another successful strategy we adopted in the iCardia4HF [16] study was to send weekly text messages to all participants reminding them to sync. A feature implemented in the iCardia platform allowed us to send an automated reminder to selected participants when their last sync time exceeds a prespecified number of days decided by the research team (eg, 6 days).

A less common issue that impacted our ability to collect Fitbit data from several participants for days, especially in the ACTION [14] and BAILA TECH [12,13] studies, was a firmware update. During this update programmed by Fitbit, several participants were not able to perform the update by themselves and were also not able to sync their trackers with the mobile app. One important issue observed in BAILA TECH [12,13] is that many participants did not have enough memory space in their phone to allow updates. We often reminded participants to delete files on the phone and make sure the updates were completed before it interfered with the app function. While the update did not change Fitbit features on the tracker and mobile app, it required the research teams to meet with participants to solve this issue. Although not common, researchers should be aware that such updates might occur in the middle of the study and disrupt data collection for some participants who are less adept with technology.

Frequent updates also occur on the Fitbit app, usually without prior notification to the research team. Some updates are minor, addressing software issues such as bugs. However, other updates, involving the addition or removal of behavior change features or tools, are more disruptive and may have significant implications for the design and execution of a study. For example, in the iCardia4HF study, changes to the app user interface while the study was ongoing required significant adjustments to study materials (eg, participants' manual) and participant training or study onboarding sessions. Also, the addition of new features such as motivational messages (which

now appear at the top of the dashboard every time a user accesses the Fitbit app), guided exercise programs, guided mindfulness and meditation, social media interactions, and user notifications may introduce new intervention components beyond those desired or anticipated by the research team, which may have an impact on the study design and results. It is important for researchers to understand and assess the impact that these changes may have in their study, in order to come up with a strategy to either deactivate such features when they become available (if possible to avoid contamination of their intervention) or interview participants at the end of the study to better understand which of these features were used, how often, and how they may have contributed to the outcomes of interest (eg, steps, moderate-to-vigorous PA, sedentary time, wear time, and engagement).

The interaction between participants and the Fitbit tracker presents, in general, challenges related to broken Fitbit trackers or charging cables and losing the Fitbit. Losing the Fitbit demands prompt actions from the research staff. A study that utilized Fitbit Zip to monitor PA levels during 16 weeks in 50 participants reported that 14 participants lost their Fitbits, and some participants (the exact number was not reported) had issues with charging cables [34]. The problem with broken or lost Fitbits is that, in some instances, participants do not report it to the research staff in a timely manner. Participants should receive instructions in the orientation session and in the manual about the importance of reporting a broken or lost Fitbit. As mentioned earlier, the research staff team should monitor participants' incoming data frequently, which in the case of a broken or lost Fitbit, would flag that the participant is not syncing the Fitbit.

To account for lost and broken Fitbits, the researcher should ensure at least a 10% Fitbit and charger surplus in study budgets in case replacements are needed. For example, in the ACTION study [14], 6 Fitbits broke, and 8 chargers were broken or lost and had to be replaced. In the BAILA TECH study [12,13], 1 Fitbit and 1 charger had to be replaced. As we had a surplus, the replacements were made quickly. Researchers also should anticipate that some participants may not be satisfied with the band color or material and that some bands can become discolored. Dissatisfaction with the Fitbit color or material can lead to reduced wear-time adherence. Therefore, we also suggest that when acquiring the Fitbits, a 10% surplus of bands of different colors and materials (eg, hypoallergenic, metal) should be purchased. Studies with follow-up durations ≥12 months may require a higher surplus (eg, 20%) given the short life expectancy of these devices and bands.

Monitoring adherence to wearing the Fitbit tracker and identifying epochs of non-wear time are 2 of the most critical and challenging factors for studies that rely on Fitbit data to assess changes in PA outcomes and also to deliver a personalized behavioral change intervention. A key feature that is currently missing from all Fitbit trackers is an advanced wear-time sensor to automatically detect when the device is removed. Without knowing when and for how long a study participant wore the Fitbit, it is difficult to monitor compliance and determine with certainty whether, for example, decreases in the number of steps and moderate-to-vigorous PA are due to sedentary behavior or nonadherence to wearing the Fitbit. Unlike research-grade activity tracking devices (eg, ActiGraph GT3X+), Fitbit PA trackers do not offer a capacitive touch technology or specialized software that can help researchers analyze the raw accelerometer data and identify epochs of nonwear time.

To address this challenge, in the BAILA TECH [12,13], ACTION [14], and iCardia4HF [16] trials, we developed and tested a novel method. This method captures continuous heart rate sensor data measured at 1- to 5-second intervals and uses this data to calculate minutes of nonwear time based on the absence of a heart pulse for 60 or more seconds. Data collection and calculation are done by the iCardia platform using the Fitbit Web API. This method is susceptible to errors if a study participant is not properly wearing the tracker on the wrist (eg, too loose or too tight). Although, with adequate personnel training and monitoring, this method has allowed our research team to gather critical wear-time data to determine compliance and distinguish between real and spurious sedentary time [22]. Importantly, this method can only be implemented with Fitbit activity trackers that have a heart rate sensor (some Fitbit models do not have this feature).

Regardless of the type of issue faced during the intervention period, we learned a single strategy that helped to address most of the challenges during the intervention delivery: devoting time to in-person intervention meetings or placing telephone calls when necessary to discuss potential challenges or issues associated with the use of Fitbit. Across the studies we conducted, we dedicated a portion of the in-person meeting to Fitbit-related content [9,10,20,35]. During this time, participants were free to externalize any concerns or issues related to Fitbit use, and research staff were ready to assist with possible problems or questions. We also recommend that, at these meetings, research staff should be available to address issues on an individual basis since some might need extra time to be resolved.

We display a summary of the challenges and solutions or strategies at the intervention delivery stage of the study in Table 3.

Table 3. Challenges and solutions or strategies for utilizing the Fitbit physical activity (PA) tracker in health interventions at the intervention delivery stage.

Challenges	Strategies and solutions
Interaction between participants and app	
Syncing-related issues include risk of data loss due to study participants forgetting to sync their Fitbit tracker with the mobile app, participants not knowing or remembering how to sync, Bluetooth turned off, internet turned off; login issues include logging off accidentally and forgetting credentials.	Offer written step-by-step instructions on how often and how to sync their tracker with the Fitbit mobile app to minimize data loss; send weekly text messages reminding all participants to sync their Fitbits or target only those who have not synced periodically[a]; email or text participants with step-by-step troubleshooting; contact the superuser; ask participants to write down login credentials in the manual or another place of easy access.
Fitbit updates during the intervention	Notify participants about app updates; update the participants' manual; determine how to handle the introduction of new features or behavior change techniques (eg, behavioral notifications about PA or CV[b] health) according to study needs.
Accidental deletion of Fitbit app	Contact the superuser (eg, family member or research assistant).
Interaction between participants and tracker	
Broken Fitbit tracker (battery does not charge or broken or lost charger), lost Fitbit tracker	Ensure at least a 10% Fitbit and charger surplus to the study budget in case Fitbit needs to be replaced; have participants use a Bluetooth locator app to help find the device.
Fitbit tracker aesthetics (width, color, discoloration of the band)	Offer multiple bands of different colors, sizes, and materials.
Skin irritation and rashes caused by the Fitbit tracker	Purchase and offer hypoallergenic bands to participants that report skin irritation and rashes; suggest participants remove the tracker or wear the tracker on the other wrist for at least 1 hour per day.
Adherence to Fitbit usage (Fitbit does not calculate wear time)	Choose a device with a heart rate monitoring feature to calculate an approximate wear time [24]; have research staff send messages if wear time <10 hours/day or build automated text message reminders to wear the Fitbit.

[a]Most Fitbit trackers can save up to 7 days of intraday data and up to 31 days of cumulative data in their memory without syncing. After that time, data are erased from the tracker to create space for the new data. This creates a risk for potential data loss if study participants do not sync their tracker regularly.

[b]CV: cardiovascular.

Data Management

Although Fitbit provides a vast amount of data, some critical data (eg, wear time) are either not readily available through Fitbit's Web API or not accurate. For those trackers that capture continuous heart rate, these data can be used as a proxy for wear time. Third-party authorized platforms (eg, iCardia) automatically calculate Fitbit wear time based on the incoming heart rate data. However, if using heart rate data directly from the Fitbit platform, research staff need to develop rules for determining wear time and manually identify gaps in heart rate that indicated nonwear time. This tedious and challenging process requires a considerable amount of time and is not advisable due to time requirements and increased likelihood of wrongly estimating wear time. Nonetheless, we recognize that this might be the only option due to Fitbit's limitation in offering complete wear-time data.

The process of Fitbit data collection and management is continuous throughout the study and requires the use of several data exportation and organization practices by research staff members. Ideally, we recommend that data should be exported from the third-party platforms, researchers' platform, or Fitbit's website at least once a week. Data are exported in a comma-delimited format (.csv) and should be appropriately organized by research staff according to the study's needs. In our experience with the BAILA TECH [12,13] and ACTION [14] studies, each week of data has a spreadsheet with all the variables of interest. This spreadsheet is comprised of daily data (ie, wear time; time spent in light, moderate, vigorous, and moderate-to-vigorous PA; and sedentary behavior) for each participant, as well as an average of each variable for each participant for that week and a grand average of all participants for that week. Importantly, the spreadsheet needs to determine valid days of data according to the wear-time rules previously adopted. Days deemed as nonvalid should not be included in participants' average, following the standard practice of studies measuring PA with accelerometry [36]. For example, if in a given week, the participant had 5 valid days and 2 nonvalid days of data, the average should be calculated based on the 5 valid days. Once the intervention data collection with Fitbit is complete, a new master spreadsheet should be created with participants' average of all the periods Fitbit was collecting data.

We display a summary of the challenges and solutions or strategies at the data management stage of the study in Table 4.

Table 4. Challenges and solutions or strategies for utilizing the Fitbit physical activity (PA) tracker in health interventions at the data management stage.

Challenges	Strategies and solutions
Determining Fitbit wear time	
Individualized manual approach (not recommended for large studies)	Review participants' daily heart rate data through the Fitbit platform to identify gaps indicating nonwear time and develop rules for determining adherence to wearing Fitbit.
Automated approach	Utilize an existing platform or application (eg, iCardia, Fitabase) that will automatically calculate Fitbit wear time based on the incoming heart rate data or develop your own application to remotely collect the data utilizing Fitbit's Web API[a].
Accuracy of sedentary minutes calculated by Fitbit	Calculate adjusted sedentary time based on heart rate and wear-time data.
Data export	Export and organize participant data weekly and before study closeout.

[a]API: application programming interface.

Study Closeout

Once the intervention is concluded, Fitbit data collection should cease. There are several actions that research staff should take to guarantee Fitbit data are not collected after the end of the study, which would violate institutional review board and HIPAA regulations. If the researcher is using a third-party platform, research staff should access each participants' Fitbit account and revoke access to the third-party platform. Importantly, this action will not affect the ability to export data that were already shared with the platform, and it will only stop the sharing of incoming data. If participants are allowed to keep the Fitbit after the end of the study, it is necessary to replace the email associated with the Fitbit account with the participants' personal emails. Research staff should provide orientation (eg, step-by-step document on how to replace the email) to study participants on how to dissociate the study's email account with the Fitbit account, or the research team can do it for participants to ensure that the email is dissociated. For example, in the ACTION [14], Virtual Coach, and BAILA TECH [12,13] studies, we developed a guide on how to dissociate the study account email with the Fitbit account. Participants should be given a time period to replace the email (eg, 1 month). In case participants do not replace the email within the time period, the research team should delete the Fitbit account permanently.

We display a summary of the challenges and solutions or strategies at the study closeout stage of the study in Table 5.

Table 5. Challenges and solutions or strategies for utilizing the Fitbit physical activity (PA) tracker in health interventions at the study closeout stage.

Challenges	Strategies and solutions
Ensuring IRB[a] compliance with data collection	If using a third-party platform, revoke access on Fitbit's website at the end of the study, dissociate the study email on the Fitbit account, create step-by-step instructions on how to dissociate the study email account from the Fitbit account; if not dissociated, delete the Fitbit account; delete the study email account; follow IRB guidelines for data destruction.

[a]IRB: institutional review board.

Discussion

In this paper, we described a number of key challenges associated with the use of Fitbit as an intervention or PA measurement tool. Difficulties with the use of Fitbit in research are encountered throughout the entire research process. Researchers should be prepared to address challenges and issues in a timely fashion to ensure that the Fitbit effectively assists participants and researchers in achieving the research goals. Most of the challenges can be addressed with proper staff training and a complete manual of operations.

We conceived and developed this paper to facilitate the understanding, uptake, and utilization of Fitbit as an mHealth tool to promote behavior change. While we aimed to present a series of challenges that researchers might encounter when using Fitbit in their research along with strategies to address these, we acknowledge that these strategies are not definitive. Their feasibility and effectiveness might vary depending on several factors (eg, funding, research staff experience and size, and research goals). Further, while our experience was with Fitbit PA trackers, many of the challenges and proposed solutions may be considered when using other commercial PA trackers, yet it was beyond the scope of this paper to make any direct comparison with them or attempt to explain how these challenges and solutions may apply to other PA trackers. We hope that our shared experiences will act as a guideline to help researchers to make conscious choices on the use of Fitbit in their research, as well as assist with more effective and intentional use of Fitbit in the research realm.

Acknowledgments

We thank Susan Vega, the Pilsen Satellite Senior Center at Casa Maravilla, Nida Shah, Phyllis P Pelt, Liete de Fatima, Marcela Garcia, Jacqueline Guzman, Stephanie Gomez, Susana Galarza, and Lauren Filkestein. We acknowledge the financial support to IGM from the CAPES Foundation, Ministry of Education of Brazil.

The BAILA TECH study was funded by the mHealth Innovation Lab and the Department of Kinesiology and Nutrition at the University of Illinois at Chicago. ACTION study was funded by NHLBI K01 HL 133370 and by the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant UL1TR002003. The iCardia4HF study was funded by the National Institute on Aging through the Midwest Roybal Center for Health Promotion and Translation award P30 AG022849. The iCardia platform has received funding from the following National Institutes of Health grants: R01 NR017635, R21 NR018281, R33 HL139454, K01 HL133370, K23 HL130524. The Virtual Coach Study was funded by the National Science Foundation through Smart and Connected Health 1838770 and Early Concept Grants for Exploratory Research 1650900.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Fitbit manual from ACTION trial.
[PDF File (Adobe PDF File), 137 KB - mhealth_v9i3e25289_app1.pdf]

References

1. Piercy KL, Troiano RP, Ballard RM, Carlson SA, Fulton JE, Galuska DA, et al. The Physical Activity Guidelines for Americans. JAMA 2018 Nov 20;320(19):2020-2028. [doi: 10.1001/jama.2018.14854] [Medline: 30418471]
2. Clarke TC, Schiller JS, Norris T. Early release of selected estimates based on data from January-June 2017 National Health Interview Survey. U.S. Department of Health and Human Services. 2017 Dec. URL: https://www.cdc.gov/nchs/data/nhis/earlyrelease/earlyrelease201712.pdf [accessed 2021-02-03]
3. Guthold R, Stevens G, Riley L, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. The Lancet Global Health 2018 Oct;6(10):e1077-e1086. [doi: 10.1016/s2214-109x(18)30357-7]
4. Number of wearable device users in the United States from 2014 to 2022. Statista. 2020. URL: https://www.statista.com/statistics/543070/number-of-wearable-users-in-the-us/ [accessed 2021-02-03]
5. Emery E, Woodhead E, Molinari V, Hunt MG. Treatment adherence in late-life. In: Lichtenberg PA, editor. Handbook of Assessment in Clinical Gerontology, 2nd ed. Cambridge, MA: Academic Press; 2010:427-457.
6. Arvidsson D, Fridolfsson J, Börjesson M. Measurement of physical activity in clinical practice using accelerometers. J Intern Med 2019 Aug 16;286(2):137-153. [doi: 10.1111/joim.12908] [Medline: 30993807]
7. Hickey AM, Freedson PS. Utility of Consumer Physical Activity Trackers as an Intervention Tool in Cardiovascular Disease Prevention and Treatment. Prog Cardiovasc Dis 2016 May;58(6):613-619. [doi: 10.1016/j.pcad.2016.02.006] [Medline: 26943981]
8. Gupta L. Fitbit's Unit Sales Shipments (2016-2021). Business Quant. 2020. URL: https://businessquant.com/fitbit-unit-sales-shipments [accessed 2021-02-03]
9. Earwear and Wristbands Drive First Quarter Growth in the Worldwide Wearables Market, Says IDC. IDC. 2020 May 28. URL: https://www.idc.com/getdoc.jsp?containerId=prUS46432620 [accessed 2021-02-03]
10. 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]
11. Fitbit Research Library. Fitabase. 2021. URL: https://www.fitabase.com/research-library/ [accessed 2021-02-03]
12. Marques IG, Kitsiou S, Gerber BS, Buchholz SW, Bustamante EE, Marquez DX. Feasibility of a Latin Dance Program with mHealth for Middle-Aged and Older Latinxs (BAILA TECH). Translational Journal of the ACSM 2021;6(1):e000143. [doi: 10.1249/TJX.0000000000000143]
13. Marques IG, Marquez DX. Feasibility of a mobile health-infused dance program for middle-aged and older Latinos. Annals of Behavioral Medicine 2018;52:S216.
14. Nyenhuis SM, Balbim GM, Dixit D, Kitsiou S, Sharp LK. Physical activity monitor wear-time among low-income minority women in a randomized clinical trial. Annals of Behavioral Medicine 2020:S216.
15. Nabulsi N, Smart M, Kuo T, Gupta I, Ziebart B, Di Eugenio B, et al. PAM7 Developing an Automated Virtual Walking Coach for Underserved, Sedentary Patients in Primary Care: Analysis of Pilot Data. Value in Health 2020 May;23:S11. [doi: 10.1016/j.jval.2020.04.060]

16. Kitsiou S. Patient-centered Mobile Health Intervention to Improve Self-care in Patients With Chronic Heart Failure (iCardia4HF). U.S. National Library of Medicine: ClinicalTrials.gov. 2018. URL: https://clinicaltrials.gov/ct2/show/NCT03642275?term=NCT03642275&draw=2&rank=1 [accessed 2021-02-03]
17. Marquez DX, Wilbur J, Hughes SL, Berbaum ML, Wilson RS, Buchner DM, et al. B.A.I.L.A. - a Latin dance randomized controlled trial for older Spanish-speaking Latinos: rationale, design, and methods. Contemp Clin Trials 2014 Jul;38(2):397-408 [FREE Full text] [doi: 10.1016/j.cct.2014.06.012] [Medline: 24969395]
18. Wilbur J, Miller AM, Fogg L, McDevitt J, Castro CM, Schoeny ME, et al. Randomized Clinical Trial of the Women's Lifestyle Physical Activity Program for African-American Women: 24- and 48-Week Outcomes. Am J Health Promot 2016 May 13;30(5):335-345. [doi: 10.1177/0890117116646342] [Medline: 27404642]
19. Juniper E, Guyatt G, Cox F, Ferrie P, King D. Development and validation of the Mini Asthma Quality of Life Questionnaire. Eur Respir J 1999 Jul;14(1):32-38 [FREE Full text] [doi: 10.1034/j.1399-3003.1999.14a08.x] [Medline: 10489826]
20. Riegel B, Dickson V, Faulkner KM. The Situation-Specific Theory of Heart Failure Self-Care: Revised and Updated. J Cardiovasc Nurs 2016;31(3):226-235. [doi: 10.1097/JCN.0000000000000244] [Medline: 25774844]
21. Nyenhuis SM, Shah N, Ma J, Marquez DX, Wilbur J, Cattamanchi A, et al. Identifying barriers to physical activity among African American women with asthma. Cogent Medicine 2019 Feb 27;6(1):1582399. [doi: 10.1080/2331205x.2019.1582399]
22. Nyenhuis S, Balbim G, Ma J, Marquez D, Wilbur J, Sharp L, et al. A Walking Intervention Supplemented With Mobile Health Technology in Low-Active Urban African American Women With Asthma: Proof-of-Concept Study. JMIR Form Res 2020 Mar 11;4(3):e13900 [FREE Full text] [doi: 10.2196/13900] [Medline: 32159520]
23. Kitsiou S, Nieto A, Mishra A, Gerber BS, Stamos T. Smartphone Ownership and Use of Mobile Health Technologies for Self-Management of Heart Failure. 2018 Presented at: Annual Symposium of the American Medical Informatics Association; November 3-7, 2018; San Francisco, CA.
24. Kitsiou S, Thomas M, Marai G, Maglaveras N, Kondos G, Arena R, et al. Development of an innovative mHealth platform for remote physical activity monitoring and health coaching of cardiac rehabilitation patients. 2017 Presented at: 2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI); February 16-19, 2017; Orlando, FL. [doi: 10.1109/bhi.2017.7897223]
25. Buchholz SW, Wilbur J, Halloway S, Schoeny M, Johnson T, Vispute S, et al. Study protocol for a sequential multiple assignment randomized trial (SMART) to improve physical activity in employed women. Contemp Clin Trials 2020 Feb;89:105921. [doi: 10.1016/j.cct.2019.105921] [Medline: 31899371]
26. Prendergast H, Lara B, Khosla S, Kitsiou S, Del Rios M, Durazo-Arvizu R, et al. Preliminary Data from a Randomized Controlled Trial for a Hypertension Education and Empowerment Intervention (TOUCHED) in an Urban, Academic Emergency Department: Opportunities in the era of COVID-19. Journal of the National Medical Association 2020 Oct;112(5):S16-S17. [doi: 10.1016/j.jnma.2020.09.039]
27. Kitsiou S, Gerber B, Buchholz S, Kansal M, Chen J, Nawrocki L, et al. Development and Pilot Testing of a Tailored Text Messaging Intervention to Improve Self-efficacy and Health Beliefs About Self-care in People with Heart Failure. Journal of Cardiac Failure 2020 Oct;26(10):S89. [doi: 10.1016/j.cardfail.2020.09.260]
28. Fitabase. 2021. URL: https://www.fitabase.com/ [accessed 2021-02-03]
29. 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]
30. Find the product that's right for you. Fitbit. 2021. URL: https://www.fitbit.com/global/us/products/find-your-fit [accessed 2021-02-03]
31. Marques IG, Cantoral J, Guzman J, Balbim GM, Marquez DX. Using a physical activity wearable tracker: Older Latinos' perceptions. Innov Aging 2018;2(Suppl 1):681.
32. Fitbit. 2021. URL: https://www.fitbit.com/us/home [accessed 2021-02-03]
33. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
34. Harrison D, Marshall P, Berthouze N, Bird J. Problems related to running longitudinal studies with commercial devices. 2014 Presented at: UBICOMP '14 ADJUNCT; September 13-17, 2014; Seattle, WA. [doi: 10.1145/2638728.2641320]
35. Gupta I, Di EB, Ziebart B, Liu B, Gerber B, Sharp L. Modeling health coaching dialogues for behavioral goal extraction. 2020 Presented at: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); November 18-21, 2019; San Diego, CA. [doi: 10.1109/bibm47256.2019.8983138]
36. 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]

Abbreviations

API: application programming interface
HF: heart failure
HIPAA: Health Insurance Portability and Accountability Act
mHealth: mobile health

PA: physical activity

Edited by G Eysenbach; submitted 27.10.20; peer-reviewed by M Johansson, M Nichols; comments to author 17.11.20; revised version received 11.01.21; accepted 20.01.21; published 01.03.21.

Please cite as:
Balbim GM, Marques IG, Marquez DX, Patel D, Sharp LK, Kitsiou S, Nyenhuis SM
Using Fitbit as an mHealth Intervention Tool to Promote Physical Activity: Potential Challenges and Solutions
JMIR Mhealth Uhealth 2021;9(3):e25289
URL: https://mhealth.jmir.org/2021/3/e25289
doi:10.2196/25289
PMID:33646135

Corrigenda and Addenda

Correction: A Self-Help App for Syrian Refugees With Posttraumatic Stress (Sanadak): Randomized Controlled Trial

Susanne Röhr[1,2*], PhD; Franziska U Jung[1*], PhD; Alexander Pabst[1], PhD; Thomas Grochtdreis[3], PhD; Judith Dams[3], PhD; Michaela Nagl[4], PhD; Anna Renner[4], MA; Rahel Hoffmann[4], MSc; Hans-Helmut König[3], MD, PhD; Anette Kersting[4], MD, PhD; Steffi G Riedel-Heller[1], MPH, MD, PhD

[1]Institute of Social Medicine, Occupational Health and Public Health, Medical Faculty, University of Leipzig, Leipzig, Germany

[2]Global Brain Health Institute, Trinity College Dublin, Dublin, Ireland

[3]Department of Health Economics and Health Services Research, Hamburg Center for Health Economics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[4]Department of Psychosomatic Medicine and Psychotherapy, University Medical Center Leipzig, Leipzig, Germany

[*]these authors contributed equally

Corresponding Author:
Susanne Röhr, PhD
Institute of Social Medicine, Occupational Health and Public Health
Medical Faculty, University of Leipzig
Philipp-Rosenthal-Strasse 55
Leipzig
Germany
Phone: 49 341 97 24568
Email: susanne.roehr@medizin.uni-leipzig.de

Related Article:

Correction of: https://mhealth.jmir.org/2021/1/e24807/

(JMIR Mhealth Uhealth 2021;9(3):e28336) doi:10.2196/28336

In "A Self-Help App for Syrian Refugees With Posttraumatic Stress (Sanadak): Randomized Controlled Trial" (JMIR Mhealth Uhealth 2021;9(1):e24807) one error was noted.

In Table 4, in the row "ED-5D-5L" and sub-row "4 weeks", under the final column "95% CI", the correct value was not rendered and appeared only as "t". This value has been corrected to "–0.189 to 0.556".

The correction will appear in the online version of the paper on the JMIR Publications website on March 5, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 01.03.21; this is a non–peer-reviewed article;accepted 03.03.21; published 05.03.21.

Please cite as:
Röhr S, Jung FU, Pabst A, Grochtdreis T, Dams J, Nagl M, Renner A, Hoffmann R, König HH, Kersting A, Riedel-Heller SG
Correction: A Self-Help App for Syrian Refugees With Posttraumatic Stress (Sanadak): Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(3):e28336
URL: https://mhealth.jmir.org/2021/3/e28336
doi:10.2196/28336
PMID:33667178

Original Paper

Tracking and Monitoring Mood Stability of Patients With Major Depressive Disorder by Machine Learning Models Using Passive Digital Data: Prospective Naturalistic Multicenter Study

Ran Bai[1,2*], MS; Le Xiao[3*], PhD, MD; Yu Guo[4], MEng; Xuequan Zhu[3], MA; Nanxi Li[3], MD; Yashen Wang[2], PhD; Qinqin Chen[2], PhD; Lei Feng[3], PhD, MD; Yinghua Wang[2], PhD; Xiangyi Yu[2], MS; Chunxue Wang[5], MD, PhD; Yongdong Hu[6], MD, PhD; Zhandong Liu[7], MD; Haiyong Xie[1,2], PhD; Gang Wang[1,3], PhD, MD

[1]Advanced Innovation Center for Human Brain Protection, Capital Medical University, Beijing, China

[2]National Engineering Laboratory for Risk Perception and Prevention, Beijing, China

[3]The National Clinical Research Center for Mental Disorders, Beijing Anding Hospital, Capital Medical University, Beijing, China

[4]Beijing University of Posts and Telecommunications, Beijing, China

[5]Department of Neuropsychiatry and Clinical Neurology, Beijing Tiantan Hospital, Capital Medical University, Beijing, China

[6]Department of Psychological Medicine, Beijing Chao-Yang Hospital, Capital Medical University, Beijing, China

[7]Department of Neurology, Medical Health Center, Beijing Friendship Hospital, Capital Medical University, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Gang Wang, PhD, MD
The National Clinical Research Center for Mental Disorders
Beijing Anding Hospital
Capital Medical University
No 5 Ankang Lane, Xicheng District
Beijing, 100088
China
Phone: 86 13466604224
Fax: 86 1058303289
Email: gangwangdoc@ccmu.edu.cn

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2021/6/e30540

Abstract

Background: Major depressive disorder (MDD) is a common mental illness characterized by persistent sadness and a loss of interest in activities. Using smartphones and wearable devices to monitor the mental condition of patients with MDD has been examined in several studies. However, few studies have used passively collected data to monitor mood changes over time.

Objective: The aim of this study is to examine the feasibility of monitoring mood status and stability of patients with MDD using machine learning models trained by passively collected data, including phone use data, sleep data, and step count data.

Methods: We constructed 950 data samples representing time spans during three consecutive Patient Health Questionnaire-9 assessments. Each data sample was labeled as Steady or Mood Swing, with subgroups Steady-remission, Steady-depressed, Mood Swing-drastic, and Mood Swing-moderate based on patients' Patient Health Questionnaire-9 scores from three visits. A total of 252 features were extracted, and 4 feature selection models were applied; 6 different combinations of types of data were experimented with using 6 different machine learning models.

Results: A total of 334 participants with MDD were enrolled in this study. The highest average accuracy of classification between Steady and Mood Swing was 76.67% (SD 8.47%) and that of recall was 90.44% (SD 6.93%), with features from all types of data being used. Among the 6 combinations of types of data we experimented with, the overall best combination was using call logs, sleep data, step count data, and heart rate data. The accuracies of predicting between Steady-remission and Mood Swing-drastic, Steady-remission and Mood Swing-moderate, and Steady-depressed and Mood Swing-drastic were over 80%,

and the accuracy of predicting between Steady-depressed and Mood Swing-moderate and the overall Steady to Mood Swing classification accuracy were over 75%. Comparing all 6 aforementioned combinations, we found that the overall prediction accuracies between Steady-remission and Mood Swing (drastic and moderate) are better than those between Steady-depressed and Mood Swing (drastic and moderate).

Conclusions: Our proposed method could be used to monitor mood changes in patients with MDD with promising accuracy by using passively collected data, which can be used as a reference by doctors for adjusting treatment plans or for warning patients and their guardians of a relapse.

Trial Registration: Chinese Clinical Trial Registry ChiCTR1900021461; http://www.chictr.org.cn/showprojen.aspx?proj=36173

(JMIR Mhealth Uhealth 2021;9(3):e24365) doi:10.2196/24365

KEYWORDS

digital phenotype; major depressive disorder; machine learning; mobile phone

Introduction

Depression is a common mental illness characterized by persistent sadness and a loss of interest in activities that people normally enjoy, accompanied by an inability to carry out daily activities for 14 days or longer [1]. The latest estimates from the World Health Organization show that more than 300 million people are now living with depression, and it has increased by more than 18% between 2005 and 2015. Treatment of major depressive disorder (MDD) usually spans a long period (no less than 6 months). Receiving continuous and long-term maintenance treatment could reduce or even prevent relapse. It is essential for doctors to monitor patients' condition and symptoms to provide appropriate treatment. However, it is impossible for doctors to track the patients' condition every day as patients revisit their doctors twice a month in an ideal case. Besides, it is not easy for patients to provide a precise description of their conditions for the past several weeks; sometimes, the answer could be as vague as an *OK*.

This study analyzed daily phone usage data, sleep data, and step count data of patients with MDD and their self-evaluated mood scores. According to a study on smartphone ownership across countries, of the top 20 countries reported, an average of 73.45% (SD 10.79%) of adults own a smartphone [2]. According to the China Netcasting Services Association [3], the average time people spend on mobile internet using their smartphones is 341.2 minutes per day in China. With the rapid evolution of smartphone and wearable device technologies, many internet-based mental health services have emerged. Many researchers are focusing on using smartphone usage data to infer mood [4-8]. Sleep and sports data collected by mobile sensors have also been studied by researchers as an inference of mood [9-13]. Jacobson et al [14] used movement and light data to assess depression severity. Cho et al [15] predicted the mood state of patients with MDD in the next 3 days using passively collected data from smartphones. Merikangas et al [16] examined the association among motor activity, energy, mood, and sleep in adults with mental disorders. Cao et al [17] used smartphone-based self-reports, parent evaluations, and passive phone sensor data to monitor depressive symptoms of adolescent patients with MDD. Canzian et al [18] investigated the correlation between patterns of human mobility and emotional states of depressive patients using GPS data collected from smartphones.

When reviewing works on mental state monitoring and predicting, we found that there are 2 major approaches: (1) training a generic model using all data collected and (2) building a personalized model for each patient. During data preprocessing, we observed differences in phone usage routines among patients. Owing to the nature of Patient Health Questionnaire-9 (PHQ-9), which reflects a patient's mental state for the past week, there were limited data samples for each patient to build a personalized model. To eliminate individual differences between patients, we examined the correlation between the change in phone usage routine, sleep data, and step count and the change in the patient's level of depression.

The main objective of this study is to examine the feasibility and technical foundation of monitoring variations in depression levels in patients with MDD during a period based on the amount of variation in smartphone usage data, sleep data, and step count data. We then analyzed different models trained by data to determine which types of behaviors were most affected by the change in their depression level.

Methods

Smartphone-Based Depression App Design

We designed an app called Mood Mirror to track and record patients' daily activities and mood (Figures 1 and 2). The goal was to collect phone usage data and physical data passively with minimal human action. Owing to the limitations of access to app usage on the iOS platform, our Mood Mirror app only supported the Android platform. The app requires users to wear a wristband that we provided to collect sleep, heart rate, and step count data.

The Mood Mirror app consists of 2 main parts: self-evaluation of mood condition and data collection. The app sends a notification to the user every day at 8 PM to use the Visual Analog Scale (VAS) to evaluate their mood of the day on a scale of −3 to 3, with −3 indicating sadness and 3 indicating happiness (Figure 3). The app also provides multiple self-rating tools such as PHQ-9 and Generalized Anxiety Disorder-7 (Figure 4). Users could use these tools to evaluate their mental state anytime. Meanwhile, with users' consents, the Mood Mirror app runs in the background to collect phone usage data, including call logs, text message logs, app usage logs, GPS, and screen on and off status. These phone usage data would be uploaded instantly to our server. In addition, the app is able to

connect with the wristband that is provided via Bluetooth. The data collected by the wristband would first be stored locally and uploaded to our server when the user connects with the wristband using the Mood Mirror app. The Mood Mirror app also allows users to record their medication prescriptions and side effects to keep track of their conditionsAll patients provided written informed consent to participate in the study. Users are able to track their mood variation history, sleep data, and step count via the Mood Mirror app. The Mood Mirror app would send notifications to remind users to keep recording their mood if the app was not used for more than 3 days.

In this study, we selected Mi Band 2 (Xiaomi Corporation), a top-selling wristband model that was sold to millions in China at the time. According to the product description, the data collected were calibrated in their research and development laboratory, and their algorithms of sleep and sports have been widely accepted.

To collect phone usage data that could reflect a subject's real daily routine, subjects were asked to install the Mood Mirror app on their own phone. The app was tested on more than 20 different models for sale at the time from top-selling brands such as HUAWEI, Xiaomi, and OPPO and had also been tested on different Android operating systems for its compatibility.

Figure 1. Home screen of the Mood Mirror app.

Figure 2. Screenshot of the menu page.

Figure 3. Screenshot of filling the Visual Analog Scale.

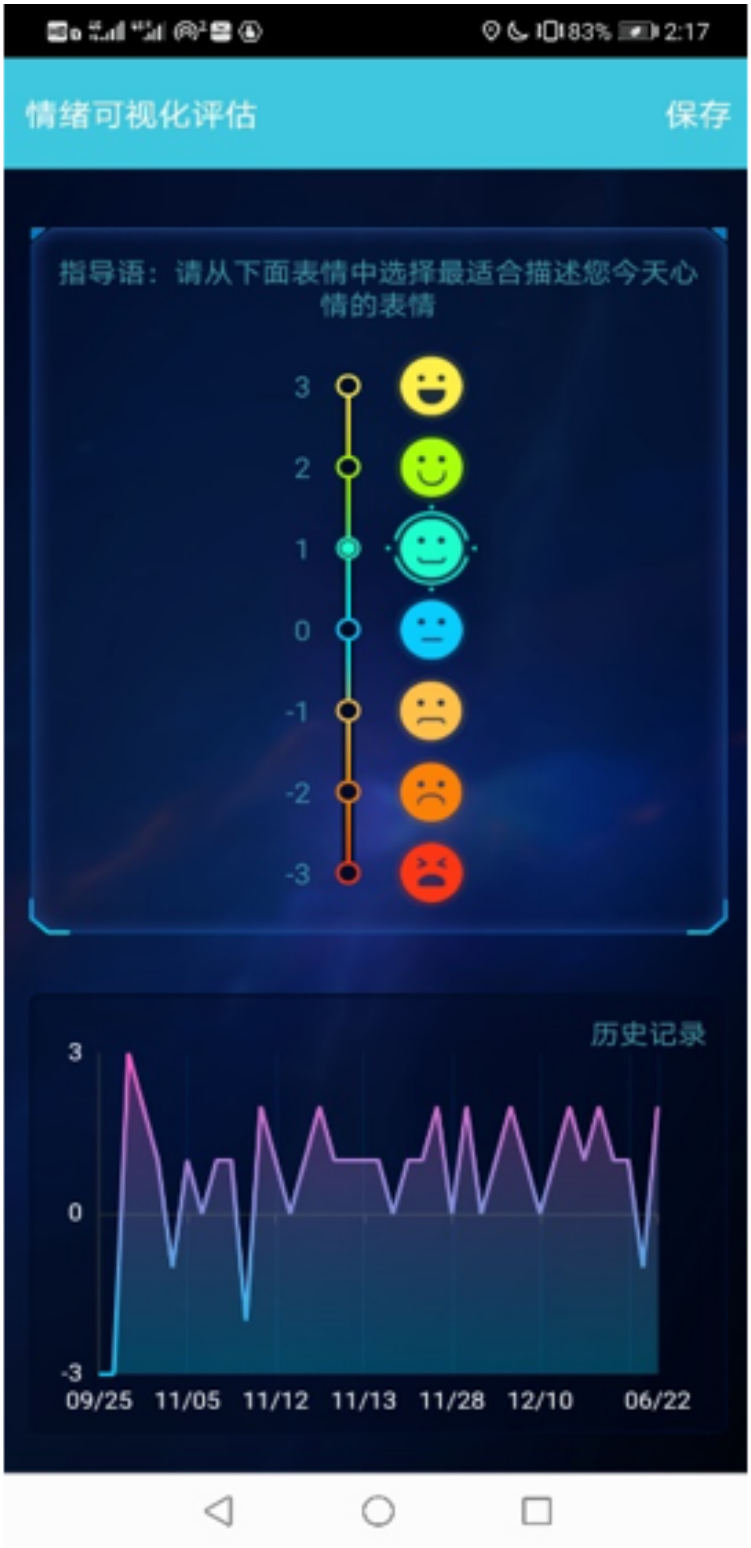


Figure 4. Screenshot of filling Patient Health Questionnaire-9.

Study Design

This was a multisite, noninterventional prospective study. The study was conducted at 4 psychiatric hospitals or units in general hospitals in Beijing, China. The protocol was approved by the Independent Medical of Ethics Committee Board of Beijing Anding Hospital and the other 3 sites (ethical approval no. 2018-119-201917FS-2). All patients provided written informed consents to participate in the study.

The study was designed to establish a correlation between clinician rating scales, self-rating scales, and passive collected phone usage measures for patients with depression. There were 4 types of data being collected:

1. Physician rating scales, including the Hamilton Depression Rating Scale, were performed by psychiatrists at each visit.
2. Self-rating scales, including PHQ-9, were performed by participants biweekly via the Mood Mirror app.

3. Daily immediate mood was recorded by participants using the VAS via the Mood Mirror app.
4. Phone usage data, including call logs, text message logs, app usage logs, GPS, and screen on and off status, were analyzed.
5. Wristband data, including sleep data, step count, and heart rate, were analyzed.

The study lasted for 12 weeks, and all participants were asked to check in with their doctors and complete the self-rating scales at weeks 0, 2, 4, 8, and 12. There was no restriction to their treatment.

All participants were explained about the study, the design of the app, and the types of data being collected by it. Each participant was then instructed to install the Mood Mirror app on his or her personal smartphone and given a wristband. Participants would connect the wristband to the app and allow the app to gain access to certain data under the assistance of a research assistant and complete self-rating scales.

During the follow-up visits, all participants were asked to record their mood status daily and complete PHQ-9 biweekly via the Mood Mirror app.

Participants

All participants were recruited from outpatient clinics at 4 sites in Beijing from February 2019 to April 2020. Participants were outpatients aged 18 to 60 years and had been diagnosed with MDD according to *theDiagnostic and Statistical Manual of Mental Disorders, fourth edition* criteria. Participants were excluded if they had Axis I primary psychiatric diagnosis other than MDD or had a diagnosis of substance abuse. Clinicians introduced the study to patients who met the study criteria in outpatient clinics. If the patients who own an Android phone were interested, the clinician would refer the patients to the research center, and a research assistant would explain the study in detail. If the patients agreed to participate in the study, the research assistant would ask them to sign an informed consent form and help with the app and wristband setups. Participants received ¥100 (US $15.5) for each follow-up visit.

Data Preprocessing and Feature Extraction

Data Preprocessing

The focus of this study is to monitor mood changes in patients with depression. To do so, the data needed to be resampled and labeled.

For each patient, every 3 consecutive PHQ-9 results and the data collected between the first and the last PHQ-9 evaluation day would be treated as 1 data sample. The data were then divided into 2 parts: (1) data collected between the first and second PHQ-9 evaluation day and (2) data collected between the second and third PHQ-9 evaluation day. These 2 parts are called PHQ-9 periods (Figure 5). As participants were allowed to complete the PHQ-9 tests and submit the scores at any time, the sample would be discarded if either period lasts less than 1 week as the PHQ-9 test mostly reflects the patient's mental state for the past week. The sample would also be discarded if there were less than 3 days of effective data in either period. On the basis of this standard, the compliance rates for phone usage, call logs, and wristband data are 65.3%, 71.1%, and 58.11%, respectively.

The samples were then labeled into 2 groups and 4 subgroups using 3 PHQ-9 results of each data sample according to the criteria shown in Table 1.

Figure 5. Example of forming a data sample. PHQ-9: Patient Health Questionnaire-9.

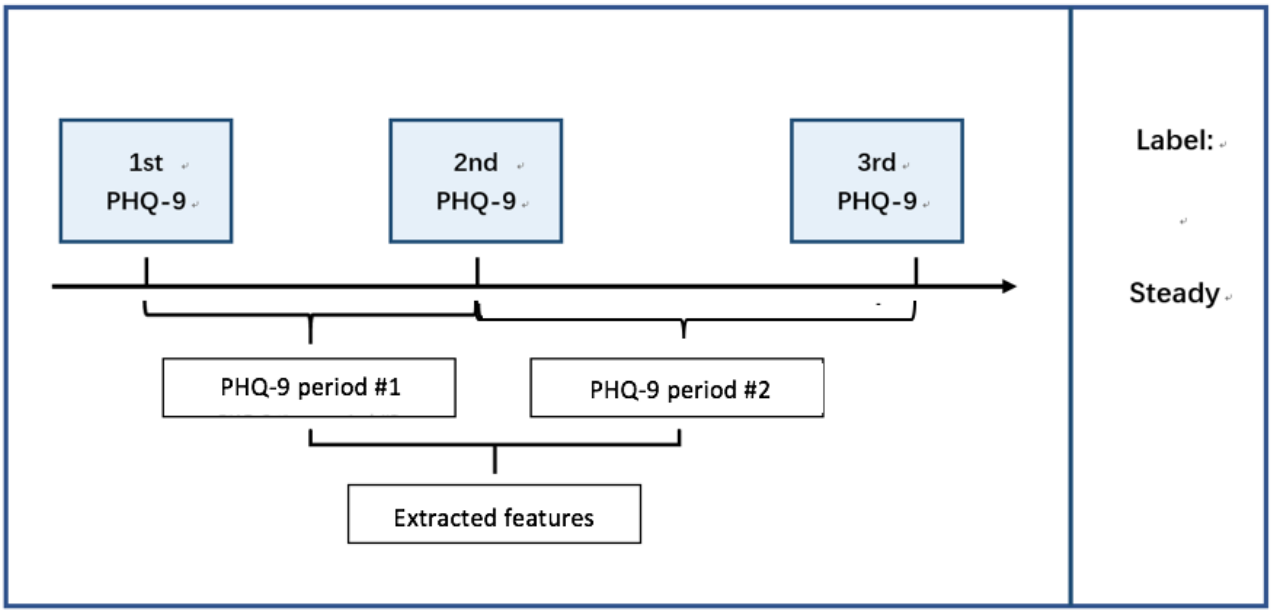


Table 1. Data labels and criteria.

Label	Criteria
Steady	
Remission	All three PHQ-9[a] results≤5
Depressed	All three PHQ-9 results≥11 and $PHQ\text{-}9_{max}-PHQ\text{-}9_{min}<5$
Swing	
Drastic	$PHQ\text{-}9_{max}-PHQ\text{-}9_{min}\geq10$
Moderate	$PHQ\text{-}9_{max}-PHQ\text{-}9_{min}\geq5$

[a]PHQ-9: Patient Health Questionnaire-9.

Feature Extraction

As data collected by the smartphone and the wristband were in different forms, the features extracted were different. There were, however, certain types of collected data that were not used in the following study based on common sense judgment and the quality of collected data. For example, text message data were not used because of the popularity of the instant messaging app WeChat. People rarely send text messages using SMS, and there was a large amount of junk messages sent by merchants and service providers. The music data were not used as well; owing to technical problems, the names of the songs were mixed with lyrics and it was difficult to clean the data without human involvement. The details of each data type that were used and extracted features are explained next.

Call Logs

It is widely believed that phone call is the key feature that reflects one's status of social life. For each phone call, the type of call (incoming, outgoing, or rejected), duration, and phone number were logged. The time of the call being made (by hour), the duration of each phone call, the number of different people the phone call was made to or from, and the entropy of callers were extracted from each type of call (incoming, outgoing, and rejected) and for all phone calls during each period.

The entropy H(X) was calculated as follows:

$$H(X)=-\Sigma\ P(X)\log_2[P(X)]$$

where P(X) is the probability of the occurrence of event X.

Each caller was considered as an event, and the probability was calculated based on the number of times he or she called, was called, or was rejected.

The difference, mean value, and SD of each feature from both PHQ-9 periods were then calculated for each data sample.

Phone Usage

The overall phone usage was calculated based on the phone screen on and off status. The Mood Mirror app logged the timestamp when the smartphone was activated or locked by the user either automatically or manually. The number of times and the duration of smartphone used were calculated by screen on and off data. The average and median of phone usage duration and the average and median of the number of times of phone usage were calculated for each period. In addition, the average duration of phone usage for each period the phone was activated was calculated. The ratio of the phone usage duration in the morning (6 AM to noon) to all day phone usage duration was calculated as well as the ratio in the afternoon (noon to 6 PM) and the ratio at night (6 PM to midnight).

The difference, mean value, and SD of each feature from both PHQ-9 periods were then calculated for each data sample.

App Usage

Apps were grouped into the following 8 categories (Table 2).

For each group, the following features were calculated:

1. The average, SD, and entropy of the app usage duration.
2. The duration of app usage in the following period: midnight to 3 AM, 3 AM to 6 AM, 6 AM to 9 AM, 9 AM to noon, noon to 3 PM, 3 PM to 6 PM, 6 PM to 9 PM, 9 PM to midnight.
3. The average, SD, and entropy of the number of times of apps being used.
4. The number of times apps were used in the following period: midnight to 3 AM, 3 AM to 6 AM, 6 AM to 9 AM, 9 AM to noon, noon to 3 PM, 3 PM to 6 PM, 6 PM to 9 PM, 9 PM to midnight.

The entropy H(X) was calculated as follows:

$$H(X)=-\Sigma\ P(X)\log_2[P(X)]$$

where P(X) is the probability of the occurrence of event X.

Table 2. Apps categories and examples.

Categories	Examples
Instant messaging	WeChat, QQ
Social networking	Weibo, Zhihu, XiaoHongShu
Shopping	Taobao, JD, PinDuoDuo
Entertainment	TikTok, Bilibili, Youku, iQiyi
Music	Netease Music, QQ Music, Xiami Music
Food delivery	Meituan, Ele.me
Others	Baidu browser, Youdao Dictionary
All apps	All apps being used

Each app category was considered as an event, and the probability was calculated based on the number of times and the duration of that category of app being used.

As messaging is one of the most common ways that people are using recently to communicate with each other, the ratio of the duration of using instant messaging apps to the duration of all apps being used was calculated as a feature to partially represent one's social life.

The difference, mean value, and SD of each feature from 2 PHQ-9 periods were then calculated for each data sample.

Sleep and Step Count

The sleep and step count data were collected using a wristband. There are 4 types of wristband data: activity, light sleep, deep sleep, and not worn.

The wristband uploaded one data packet per minute, containing timestamp, data type, activity intensity, step count, and heart rate.

For sleep data, the average, median, and SD of light sleep, deep sleep, and total sleeping durations were calculated. The ratio of the light sleep duration to the total sleep duration and the ratio of the deep sleep duration to the total sleep duration were calculated as a reference of sleep quality. The time of falling into sleep and wake-up time were also used as features to estimate the user's daily routine.

For step count data, the total step count for each period was calculated. The average, median, and SD of daily step count and of the following period were calculated as well: midnight to 3 AM, 3 AM to 6 AM, 6 AM to 9 AM, 9 AM to noon, noon to 3 PM, 3 PM to 6 PM, 6 PM to 9 PM, and 9 PM to midnight.

The difference, mean value, and SD of each feature from both PHQ-9 periods were then calculated for each data sample.

Heart Rate

Heart rate data were collected using a wristband with a sampling rate of one piece of data per minute. Heart rate data were collected only when the wristband detected the user was in a light sleep mode or in a deep sleep mode.

A cosinor analysis (cosine curve fitting) was performed on heart rate data of each night. The amplitude, acrophase (peak), mesor (mean), and r-squared value (strength) were then generated from the cosine curve, and the average, median, and SD were calculated.

The difference, mean value, and SD of each feature from both PHQ-9 periods were then calculated for each data sample.

Feature Selection and Machine Learning Models

Feature Selection

With all calculated features, it was important to determine which subset of features could best describe the difference between participants who were in a steady mood and those with a mood swing. In this study, 2 different feature selection models were experimented to find a better subset of features that delivered the best accuracy and recall of classification and to avoid overfitting of data.

L1-Based Feature Selection

The L1-based feature selection method takes advantage of the fact that linear models using L1 regularization have sparse solutions. L1 regularization adds the sum of the absolute values of the coefficient as a penalty term. Owing to the inherent linear dependence on the model parameters, L1 regularization disables irrelevant features and produces sparse sets of features [19].

Tree-Based Feature Selection

The tree-based feature selection method adopts the interpretability of the tree model. The importance score of each feature is calculated, with each feature contributing to the final decision. By ranking all the importance scores, the features with lower scores contribute less to the final decision and can be removed.

Machine Learning Models

In this study, some of the most classic machine learning (ML) models were deployed to learn from the features extracted earlier and make predictions. To obtain a more accurate result, 10-fold cross-validation was performed for each subset of features of each model.

The average accuracy rate and recall rate of all 10 folds were calculated to estimate the performance of the model.

The ML models used were support vector machines (SVMs), K-nearest neighbors, decision trees, naïve Bayes, random forest, and logistic regression.

SVM

An SVM is a supervised ML model that can be used for classification. The SVM algorithm creates a line, a hyperplane, or a set of hyperplanes and maximizes the margin around it to separate data into classes.

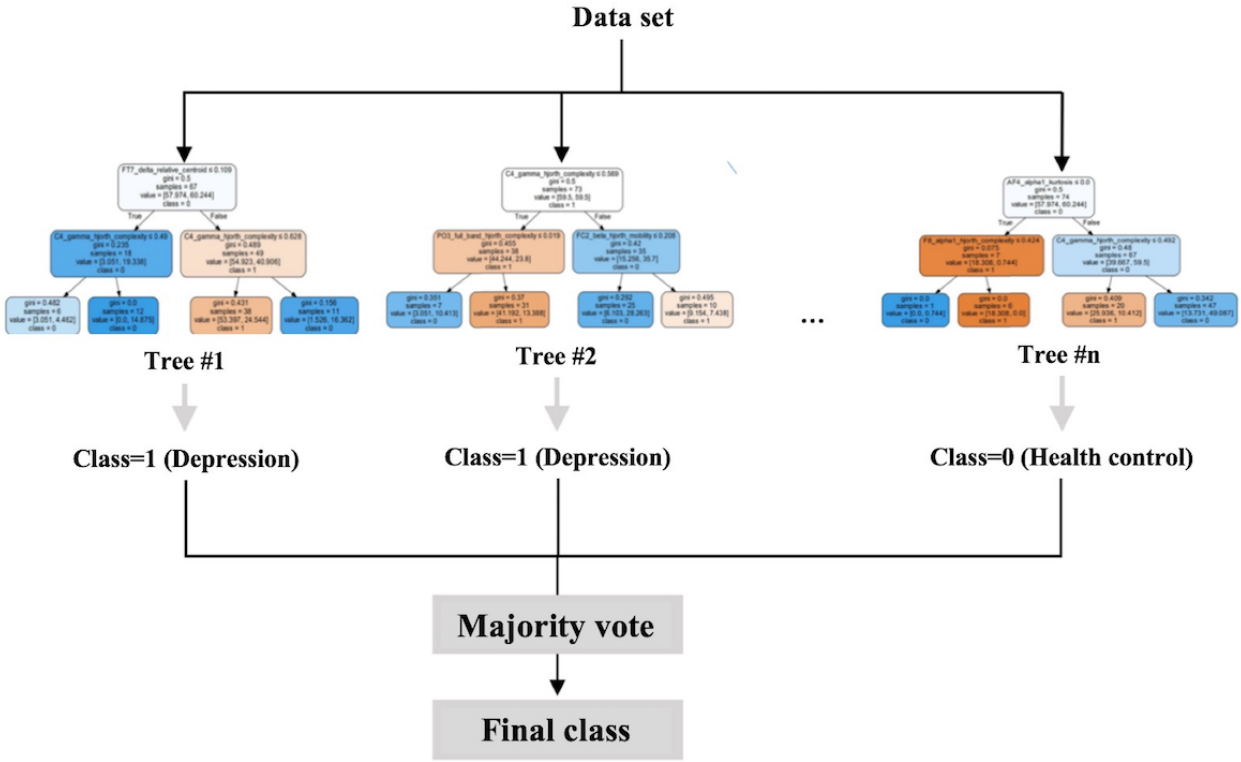
Decision Tree

A decision tree is a tree-like predictive model. In a decision tree, each interior node represents an input feature, the leaf node represents the class label, and the branches represent the decision-making progress from nodes to leaves.

Random Forest

Random forests, as shown in Figure 6, are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest [20]. It is an ensemble learning method for classification. Random forests grow many decision trees. When classifying, the input is put to each decision tree and each tree returns a classification result, and the trees *vote* for the final result. The forest then returns the final classification result with the most votes [21].

Figure 6. Mechanism of a random forest.



Results

A total of 334 participants were enrolled in this study. Owing to technical limitations and participants' different degrees of involvement, the amount of usable data samples is limited. Of 334 participants, 261 contributed 950 data samples that were suitable for analysis. As the data collection mechanisms differed between the Mood Mirror app and the wristband, there were discrepancies between the number of phone usage data samples and the number of sleep data samples. The numbers of data samples used for each model are shown in Tables 3-8.

Table 3. Classification result using selected features of phone data.

Two classes or subclasses being predicted (number of data samples)	Features selected, n	Best ML[a] model	Average percent accuracy (SD)	Average percent recall (SD)
Steady (n=144) and Swing (n=234)	4	Random forest	66.76 (4.94)	80.93 (7.72)
Steady-remission (n=25) and Swing-drastic (n=75)	36	Random forest	70.74 (6.62)	77.58 (7.12)
Steady-remission (n=25) and Swing-moderate (n=159)	7	Random forest	80.92 (5.34)	95.50 (2.30)
Steady-depressed (n=119) and Swing-drastic (n=75)	10	Decision Tree	66.18 (6.31)	65.71 (6.99)
Steady-depressed (n=119) and Swing-moderate (n=159)	34	Random forest	75.23 (3.75)	88.99 (6.00)

[a]ML: machine learning.

Table 4. Classification result using selected features of sleep data.

Two classes or subclasses being predicted (number of data samples)	Features selected, n	Best ML[a] model	Average percent accuracy (SD)	Average percent recall (SD)
Steady (n=230) and Swing (n=382)	48	Random forest	72.70 (4.74)	90.80 (3.92)
Steady-remission (n=88) and Swing-drastic (n=124)	44	Random forest	77.34 (7.50)	90.61 (6.23)
Steady-remission (n=88) and Swing-moderate (n=258)	17	Random forest	84.46 (5.94)	97.38 (2.95)
Steady-depressed (n=142) and Swing-drastic (n=124)	48	Random forest	68.87 (9.34)	67.09 (9.19)
Steady-depressed (n=142) and Swing-moderate (n=258)	5	Random forest	74.75 (5.96)	90.37 (5.18)

[a]ML: machine learning.

Table 5. Classification result using selected features of step count data.

Two classes or subclasses being predicted (number of data samples)	Features selected, n	Best ML[a] model	Average percent accuracy (SD)	Average percent recall (SD)
Steady (n=138) and Swing (n=246)	11	Random forest	69.24 (8.54)	86.97 (7.35)
Steady-remission (n=31) and Swing-drastic (n=78)	10	KNN[b]	76.09 (8.49)	96.53 (5.32)
Steady-remission (n=31) and Swing-moderate (n=168)	9	Random forest	85.42 (5.69)	99.41 (1.76)
Steady-depressed (n=107) and Swing-drastic (n=78)	8	Logistic regression	70.35 (8.57)	84.16 (11.82)
Steady-depressed (n=107) and Swing-moderate (n=168)	12	Random forest	72.33 (7.55)	84.57 (8.41)

[a]ML: machine learning.

[b]KNN: K-nearest neighbors.

Table 6. Classification result using selected features of heart rate data.

Two classes or subclasses being predicted (number of data samples)	Features selected, n	Best ML[a] model	Average percent accuracy (SD)	Average percent recall (SD)
Steady (n=80) and Swing (n=122)	20	Random forest	75.19 (8.38)	91.92 (6.71)
Steady-remission (n=18) and Swing-drastic (n=48)	9	KNN[b]	75.48 (16.53)	85.17 (15.10)
Steady-remission (n=18) and Swing-moderate (n=74)	13	KNN	82.67 (10.03)	97.64 (4.73)
Steady-depressed (n=62) and Swing-drastic (n=48)	8	Decision tree	74.55 (13.97)	73.79 (16.04)
Steady-depressed (n=62) and Swing-moderate (n=74)	18	Random forest	69.29 (13.21)	75.16 (13.96)

[a]ML: machine learning.

[b]KNN: K-nearest neighbors.

Table 7. Classification result using selected features of all data collected.

Two classes or subclasses being predicted (number of data samples)	Features selected, n	Best ML[a] model	Average percent accuracy (SD)	Average percent recall (SD)
Steady (n=79) and Swing (n=122)	75	KNN[b]	76.67 (8.47)	90.44 (6.93)
Steady-remission (n=18) and Swing-drastic (n=48)	7	Naïve Bayes	74.29 (9.27)	84.31 (10.89)
Steady-remission (n=18) and Swing-moderate (n=74)	8	KNN	80.56 (15.28)	97.08 (5.91)
Steady-depressed (n=61) and Swing-drastic (n=48)	7	Logistic regression	75.91 (13.18)	89.83 (10.34)
Steady-depressed (n=61) and Swing-moderate (n=74)	12	SVM[c]	74.73 (8.44)	83.95 (12.27)

[a]ML: machine learning.

[b]KNN: K-nearest neighbors.

[c]SVM: support vector machine.

Table 8. Classification result using selected features of call logs, sleep data, step count data, and heart rate data.

Two classes or subclasses being predicted (number of data samples)	Features selected, n	Best ML[a] model	Average percent accuracy (SD)	Average percent recall (SD)
Steady (n=79) and Swing (n=122)	37	Random forest	75.64 (5.09)	89.93 (7.26)
Steady-remission (n=18) and Swing-drastic (n=48)	8	Naïve Bayes	81.67 (15.32)	93.33 (10.41)
Steady-remission (n=18) and Swing-moderate (n=74)	7	Decision tree	80.56 (10.49)	92.88 (10.43)
Steady-depressed (n=61) and Swing-drastic (n=48)	35	Random forest	84.27 (14.36)	85.33 (15.72)
Steady-depressed (n=61) and Swing-moderate (n=74)	25	SVM[b]	77.86 (8.90)	88.99 (9.76)

[a]ML: machine learning.

[b]SVM: support vector machine.

Table 3 presents the classification results of predicting mood changes using selected features of phone data, including app usage data and call logs. It is observed that the classification between Steady-remission and Swing-moderate has the highest accuracy rate of 80.92% and recall rate of 95.50%. The classification between Steady-depressed and Swing-drastic has the lowest accuracy rate of 66.18% and recall rate of 65.71%. The classification between all Steady status samples and all Swing data samples has an accuracy rate of 66.76% and a recall rate of 80.93%.

Table 4 describes the classification results of predicting mood changes using the selected features of sleep data. The classification between Steady-remission and Swing-moderate has the highest accuracy rate (84.46%) and recall rate (97.38%). The classification between Steady-depressed and Swing-drastic has the lowest accuracy rate of 68.87% and recall rate of 67.09%. The classification between all Steady data samples and all Swing data samples has an accuracy rate of 72.70% and a recall rate of 90.80%.

The classification results of predicting mood changes using selected features of step count data show that the classification between Steady-remission and Swing-moderate has the highest accuracy rate of 85.42% and recall rate of 99.41%. The classification between all Steady data samples and all Swing data samples has the lowest accuracy rate of 69.24% and recall rate of 86.97% (Table 5).

Table 6 presents the classification results of predicting mood changes using the selected features of heart rate data. The classification between Steady-remission and Swing-moderate has the highest accuracy rate of 82.67% and recall rate of 97.64%. The classification between Steady-depressed and Swing-moderate has the lowest accuracy rate of 69.29% and recall rate of 75.16%. The classification between all Steady data samples and all Swing data samples has an accuracy rate of 75.19% and a recall rate of 91.92%.

Table 7 compares the classification results of predicting mood changes using the selected features of all data collected. The classification between Steady-remission and Swing-moderate has the highest accuracy rate of 80.56% and recall rate of 97.08%. The classification between Steady-remission and Swing-drastic has the lowest accuracy rate of 74.29% and recall rate of 84.31%. The classification between all Steady data samples and all Swing data samples has an accuracy rate of 76.67% and a recall rate of 90.44%.

The classification results of predicting mood changes using selected features of call logs, sleep data, step count data, and heart rate data show that the classification between Steady-depressed and Swing-drastic has the highest accuracy rate of 84.27% and recall rate of 85.33%. The classification between all Steady data samples and all Swing data samples has the lowest accuracy rate of 75.64% and a recall rate of 89.93% (Table 8).

Discussion

Principal Findings

To our knowledge, this study is the first to investigate the prediction of mood swings in patients with MDD by using the amount of variation in phone data, sleep data, and step count data in a period.

In this study, we calculated over hundreds of features from phone data, sleep data, and step count data and used different feature selection models to find features that could best represent the data. Multiple ML models were applied, and different combinations of types of data were examined to select the types of data to collect for future applications.

Most of the models have accuracies of more than 70%, showing promising results using passively collected phone and wristband data to predict whether patients with MDD have mood swings.

Among the 6 combinations of types of data we experimented, the overall best combination was using call logs, sleep data, step count data, and heart rate data. Accuracies of predicting between Steady-remission and Mood Swing-drastic, Steady-remission and Mood Swing-moderate, and Steady-depressed and Mood Swing-drastic are more than 80%, and accuracies of predicting between Steady-depressed and Mood Swing-moderate and the overall Steady to Mood Swing classification accuracy were over 75%. The features used in this model included the average, SD, and median of the following: sleep duration, deep sleep duration, light sleep duration, the ratio of the deep sleep duration to all-night sleep duration, the ratio of the light sleep duration to all-night sleep duration, step counts for each 3-hour period of a day, number of people called (incoming and outgoing calls), number of rejected calls, number of answered calls, and r-squared of heart rate fitted curves. We consider that the features chosen by the model reflect some of the depressive symptoms (PHQ-9) of patients with MDD: low sleep quality, reduced social interaction,

and reduced physical activity. These features are consistent with clinical phenotypes such as sleep disturbance, loss of interest, social isolation, and fatigue.

Comparing all the 6 aforementioned combinations, we found that overall prediction accuracies between Steady-remission and Mood Swing (drastic and moderate) are better than those between Steady-depressed and Mood Swing (drastic and moderate). We think that patients who continuously show depressed symptoms might have a similar behavior pattern to patients who have mood swings. On the other hand, the differences in daily behavior patterns between patients who are in remission and those who have mood swings might be more significant. This could explain why the classification accuracies between all Steady data samples and all Mood Swing data samples are lower, sometimes the lowest among all classifications, even with the largest data training set.

We found that models using features from all collected data had lower accuracies than those using features from all collected data except for app usage data (Tables 7 and 8). This might suggest that the differences in app usage behaviors are insignificant between patients who are in Steady status and those who have mood swings. Meanwhile, among the 6 combinations of types of data, models using phone data, including app usage and call logs, have the lowest overall accuracies.

Limitations and Future Work

We observed a data imbalance in our data set with a low prevalence of the Steady-remission class. As recruitment was done in the hospital outpatient department, the severity of depressive symptoms among patients was different, and there were limited data samples of patients who were in remission. The imbalance of data caused most of the models mentioned earlier to have a much higher recall rate compared with accuracy rates.

The overall data size was also limited. With a larger data set, the prediction model could be more robust. We recruited 334 participants, and all of them were asked to use the app as frequently as possible to record their mood and depression level for 12 weeks. Owing to certain restrictions on the Android system, it was difficult to keep our app running in the background 24×7 collecting data.

This study has shown the possibility of using digital phenotyping data to detect MDD patients' mood stability. We are currently working on a new version of the Mood Mirror app; with more utility functions provided and interaction designs, patients could gain more information about their current condition, which could increase patients' compliance rate and enhance both the size and quality of data. The current prediction model will be installed on this version and will provide predictions of patients' mood stability. The app would ask for patients' feedback on the prediction results. The performance of the models could be improved by a larger and more balanced data set along with the prediction results feedback.

Conclusions

This study verified the feasibility of using the amount of variation in smartphone data, sleep data, and step count data to predict whether a patient with MDD has a mood swing that should be noticed by his or her guardian and doctors. The key novelty of this study is instead of predicting the mood state of a certain point, we focus on the variation of mood over a period using the amount of variation in passive digital data. The study was limited by the imbalance of data samples and the technical constraint that the app only runs on the Android platform.

Acknowledgments

This research was funded by the Capital's Funds for Health Improvement and Research (2018-1-2121), the National Science and Technology Major Project for IND (2018ZX09201-014), Beijing Municipal Administration of Hospitals Clinical Medicine Development of Special Funding Support (XMLX201848), and Beijing Hospitals Authority Youth Program (QML20181902 and QML20181904).

Conflicts of Interest

None declared.

References

1. Depression: Let's Talk? Says Who. World Health Organization. 2017. URL: https://www.who.int/news-room/detail/30-03-2017--depression-let-s-talk-says-who-as-depression-tops-list-of-causes-of-ill-health [accessed 2017-03-30]
2. Silver L. Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally. URL: https://tinyurl.com/3tzy3467
3. 2019 China Netcasting Development Research Report. URL: http://www.cnsa.cn/home/industry/download/id/692.html
4. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, 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]
5. Asselbergs J, Ruwaard J, Ejdys M, Schrader N, Sijbrandij M, Riper H. Mobile phone-based unobtrusive ecological momentary assessment of day-to-day mood: an explorative study. J Med Internet Res 2016 Mar 29;18(3):e72 [FREE Full text] [doi: 10.2196/jmir.5505] [Medline: 27025287]

6. Rachuri KK, Musolesi M, Mascolo C, Rentfrow PJ, Longworth C, Aucinas A. EmotionSense: A Mobile Phones based Adaptive Platform for Experimental Social Psychology Research. In: The 2010 ACM Conference on Ubiquitous Computing. 2010 Sep Presented at: Ubicomp'10; September 1-9, 2010; Copenhagen Denmark p. 281-290. [doi: 10.1145/1864349.1864393]
7. Ma Y, Xu B, Bai Y, Sun G, Zhu R. Daily Mood Assessment Based on Mobile Phone Sensing. In: Ninth International Conference on Wearable and Implantable Body Sensor Networks. 2012 May 17 Presented at: BSN'12; May 9-12, 2012; London, UK. [doi: 10.1109/bsn.2012.3]
8. Bogomolov A, Lepri B, Ferron M, Pianesi F, Pentland AS. Pervasive Stress Recognition for Sustainable Living. In: International Conference on Pervasive Computing and Communication Workshops. 2014 May 15 Presented at: IEEE'14; March 24-28, 2014; Budapest, Hungary. [doi: 10.1109/percomw.2014.6815230]
9. Muaremi A, Arnrich B, Tröster G. Towards measuring stress with smartphones and wearable devices during workday and sleep. Bionanoscience 2013 May;3:172-183 [FREE Full text] [doi: 10.1007/s12668-013-0089-2] [Medline: 25530929]
10. Rabbi M, Ali S, Choudhury T, Berke E. Passive and In-situ Assessment of Mental and Physical Well-being using Mobile Sensors. In: Proceedings of the 13th international conference on Ubiquitous computing. 2011 Sep Presented at: Ubicomp'11; September, 2011; Beijing China p. 385-394 URL: http://europepmc.org/abstract/MED/25285324 [doi: 10.1145/2030112.2030164]
11. Moturu ST, Khayal I, Aharony N, Pan W, Pentland A. Using Social Sensing to Understand the Links Between Sleep, Mood, and Sociability. In: Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing. 2011 Presented at: IEEE'11; October 9-11, 2011; Boston, MA, USA. [doi: 10.1109/passat/socialcom.2011.200]
12. Sano A, Picard RW. Stress Recognition Using Wearable Sensors and Mobile Phones. In: Humaine Association Conference on Affective Computing and Intelligent Interaction. 2013 Dec 12 Presented at: ACII'13; September 2-5, 2013; Geneva, Switzerland. [doi: 10.1109/acii.2013.117]
13. 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: The 2014 ACM Conference on Ubiquitous Computing. 2014 Presented at: UbiComp'14; September 2014; Seattle Washington. [doi: 10.1145/2632048.2632054]
14. Jacobson NC, Weingarden H, Wilhelm S. Using digital phenotyping to accurately detect depression severity. J Nerv Ment Dis 2019 Oct;207(10):893-896. [doi: 10.1097/NMD.0000000000001042] [Medline: 31596769]
15. Cho C, Lee T, Kim M, In HP, Kim L, Lee H. Mood prediction of patients with mood disorders by machine learning using passive digital phenotypes based on the circadian rhythm: prospective observational cohort study. J Med Internet Res 2019 Apr 17;21(4):e11029 [FREE Full text] [doi: 10.2196/11029] [Medline: 30994461]
16. Merikangas KR, Swendsen J, Hickie IB, Cui L, Shou H, Merikangas AK, et al. Real-time mobile monitoring of the dynamic associations among motor activity, energy, mood, and sleep in adults with bipolar disorder. JAMA Psychiatry 2019 Mar 1;76(2):190-198. [doi: 10.1001/jamapsychiatry.2018.3546] [Medline: 30540352]
17. Cao J, Truong AL, Banu S, Shah AA, Sabharwal A, Moukaddam N. Tracking and predicting depressive symptoms of adolescents using smartphone-based self-reports, parental evaluations, and passive phone sensor data: development and usability study. JMIR Ment Health 2020 Jan 24;7(1):e14045 [FREE Full text] [doi: 10.2196/14045] [Medline: 32012072]
18. Canzian L, Musolesi M. Trajectories of depression: Unobtrusive Monitoring of Depressive States by means of Smartphone Mobility Traces Analysis. In: International Joint Conference on Pervasive and Ubiquitous Computing. 2015 Presented at: ACM'15; September, 2015; Osaka Japan. [doi: 10.1145/2750858.2805845]
19. Demir-Kavuk O, Kamada M, Akutsu T, Knapp E. Prediction using step-wise L1, L2 regularization and feature selection for small data sets with large number of features. BMC Bioinformatics 2011;12(1):412. [doi: 10.1186/1471-2105-12-412]
20. Breiman L. Random Forests. Mach Learn 2001 [FREE Full text] [doi: 10.1023/A:1010933404324]
21. Breiman L, Cutler A. Random Forests. Statistics. URL: https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm#giniimp [accessed 2020-12-01]

Abbreviations

MDD: major depressive disorder
ML: machine learning
PHQ-9: Patient Health Questionnaire-9
SVM: support vector machine
VAS: Visual Analog Scale

Edited by L Buis; submitted 16.09.20; peer-reviewed by J Zhang, M Asgari Mehrabadi; comments to author 30.10.20; revised version received 27.11.20; accepted 05.01.21; published 08.03.21.

Please cite as:
Bai R, Xiao L, Guo Y, Zhu X, Li N, Wang Y, Chen Q, Feng L, Wang Y, Yu X, Wang C, Hu Y, Liu Z, Xie H, Wang G
Tracking and Monitoring Mood Stability of Patients With Major Depressive Disorder by Machine Learning Models Using Passive Digital Data: Prospective Naturalistic Multicenter Study
JMIR Mhealth Uhealth 2021;9(3):e24365
URL: https://mhealth.jmir.org/2021/3/e24365
doi:10.2196/24365
PMID:33683207

Original Paper

Predicting Emotional States Using Behavioral Markers Derived From Passively Sensed Data: Data-Driven Machine Learning Approach

Emese Sükei[1], MSc; Agnes Norbury[2], PhD; M Mercedes Perez-Rodriguez[2], MD, PhD; Pablo M Olmos[1,3], PhD; Antonio Artés[1,3], PhD

[1]Signal Theory and Communications Department, Universidad Carlos III de Madrid, Leganés, Spain

[2]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[3]Gregorio Marañón Health Research Institute, Madrid, Spain

Corresponding Author:
Emese Sükei, MSc
Signal Theory and Communications Department
Universidad Carlos III de Madrid
Torres Quevedo Bldg
Av de la Universidad, 30
Leganés, 28911
Spain
Phone: 34 916248839
Email: esukei@tsc.uc3m.es

Abstract

Background: Mental health disorders affect multiple aspects of patients' lives, including mood, cognition, and behavior. eHealth and mobile health (mHealth) technologies enable rich sets of information to be collected noninvasively, representing a promising opportunity to construct behavioral markers of mental health. Combining such data with self-reported information about psychological symptoms may provide a more comprehensive and contextualized view of a patient's mental state than questionnaire data alone. However, mobile sensed data are usually noisy and incomplete, with significant amounts of missing observations. Therefore, recognizing the clinical potential of mHealth tools depends critically on developing methods to cope with such data issues.

Objective: This study aims to present a machine learning–based approach for emotional state prediction that uses passively collected data from mobile phones and wearable devices and self-reported emotions. The proposed methods must cope with high-dimensional and heterogeneous time-series data with a large percentage of missing observations.

Methods: Passively sensed behavior and self-reported emotional state data from a cohort of 943 individuals (outpatients recruited from community clinics) were available for analysis. All patients had at least 30 days' worth of naturally occurring behavior observations, including information about physical activity, geolocation, sleep, and smartphone app use. These regularly sampled but frequently missing and heterogeneous time series were analyzed with the following probabilistic latent variable models for data averaging and feature extraction: mixture model (MM) and hidden Markov model (HMM). The extracted features were then combined with a classifier to predict emotional state. A variety of classical machine learning methods and recurrent neural networks were compared. Finally, a personalized Bayesian model was proposed to improve performance by considering the individual differences in the data and applying a different classifier bias term for each patient.

Results: Probabilistic generative models proved to be good preprocessing and feature extractor tools for data with large percentages of missing observations. Models that took into account the posterior probabilities of the MM and HMM latent states outperformed those that did not by more than 20%, suggesting that the underlying behavioral patterns identified were meaningful for individuals' overall emotional state. The best performing generalized models achieved a 0.81 area under the curve of the receiver operating characteristic and 0.71 area under the precision-recall curve when predicting self-reported emotional valence from behavior in held-out test data. Moreover, the proposed personalized models demonstrated that accounting for individual differences through a simple hierarchical model can substantially improve emotional state prediction performance without relying on previous days' data.

Conclusions: These findings demonstrate the feasibility of designing machine learning models for predicting emotional states from mobile sensing data capable of dealing with heterogeneous data with large numbers of missing observations. Such models may represent valuable tools for clinicians to monitor patients' mood states.

(JMIR Mhealth Uhealth 2021;9(3):e24465) doi:10.2196/24465

KEYWORDS

mental health; affect; mobile health; mobile phone; digital phenotype; machine learning; Bayesian analysis; probabilistic models; personalized models

Introduction

Passively Sensing Behavioral Biomarkers

The subjective experience of mood is one of the most valuable sources of information about an individual's mental health [1]. Self-reported mood is a critical component of the mental status exam interview, which is to psychiatry what the physical exam is to other fields of medicine [2]. Furthermore, clinicians routinely ask questions about mood during clinical encounters. The presence of a specific mood state is a required criterion for many psychiatric diagnoses according to the Diagnostic and Statistical Manual of Mental Disorders, fifth edition (eg, depressed mood to diagnose a major depressive episode; elevated, expansive, or irritable moods to diagnose a manic episode; etc). Mood is a predictor of psychiatric outcomes, and mood changes can be a harbinger for psychiatric decompensations. Therefore, accurate monitoring of mood states is a crucial component of mental health care. For example, both valences of mood states [3] and their variability [4] have been shown to predict important outcomes, such as several binge-eating episodes in bulimia nervosa [4] and treatment adherence in patients with bipolar disorder and opioid use disorders [3,5].

Until recently, information about mood was only available to clinicians by directly questioning patients in person, either over the phone or via telepsychiatry video platforms. However, technological advances over the last few decades have allowed for real-time monitoring of patients' self-reported mood states. Smartphone-delivered ecological momentary assessment (EMA), also known as experience sampling, "assesses individuals' current experiences, behaviors, and moods as they occur in real-time and in their real-world settings" [6]. However, despite these technological advances, this form of mood state assessment relies on an individual's current level of insight and willingness and ability to interact with the EMA platform. Many psychiatric disorders cause behavioral changes that may decrease an individual's likelihood of interacting with an EMA tool (demotivation, apathy, and survey fatigue), causing missing data, not at random. Therefore, identifying objective behavioral biomarkers of mood states that can be passively sensed without patient participation is a research priority.

Through patients' mobile phones and other wearable devices, continuous sensor data can be collected in a noninvasive manner, providing valuable information about everyday activity patterns. The possibility of inferring emotional states by analyzing smartphone use data [7-9], GPS traces of movement [10,11], social media data [12], and even sound recordings [13,14] has become a growing research focus over the past decade. Such approaches can be used to analyze individuals' emotional patterns, enabling the better self-management of one's activities and behavioral choices. Moreover, for patients with mental illnesses and their caregivers and health care providers, these models could provide a means to predict mental health crises and maladaptive behavioral patterns and allow for early intervention.

Related Work

In the past few years, numerous studies have demonstrated the potential of exploiting mobile sensing data to infer users' emotional states and well-being. In an older study, LiKamWa et al [7] developed MoodScope, a statistical inference model for predicting the users' daily mood average based on the circumplex mood model [15,16], from communication history and app use patterns. They collected data from 32 participants over 2 months and reported an initial accuracy of 66%, which improved over time for personalized models.

Jaques et al [17] conducted a study using physiological signals, location, smartphone logs, and survey responses collected over a month from 206 college students to model students' happiness. They applied classical machine learning methods, such as support vector machines (SVMs), random forests (RFs), neural networks, logistic regression (LR), k-nearest neighbor, naive Bayes, and Adaboost, to perform the classification task and reported 70% accuracy.

Another study focusing on predicting college students' stress and mental health status was conducted by Sano et al [18]. They compared lasso regression and SVM with linear and radial basis function kernels for 2 classification tasks: low or high stress and low or high mental health categories. They reported over 70% accuracy and showed a significant performance increase when data from wearable sensors (such as skin conductance and temperature) were used, compared with behavioral data derived from phone sensing.

Umematsu et al [19] compared nontemporal (SVM and LR) and temporal (long short-term memory [LSTM]) machine learning methods to forecast the stress level of the upcoming day using a predefined number of days of previous data (physiological signals, mobile phone use, location, and behavioral surveys). A more recent study by Morshed et al [20], who used the StudentLife [21] and Tesserae [22] data sets, demonstrated that mood instabilities (computed from the mapping of moods on the photographic affect meter scale [23] to arousal and valence values) are predictable from features derived from passive sensor measurements.

In a large-scale study conducted by Servia-Rodríguez et al [24], the researchers used passive sensing data and self-reported moods collected for about 3 years from 18,000 users to build a predictive model for users' mood. They trained a deep neural network of stacked restricted Boltzmann machines for a 2-class classification problem (positive and negative mood). They reported above 60% prediction accuracy for weekdays and 70% for weekends.

An LSTM recurrent neural network (RNN)–based analysis, performed by Suhara et al [25], showed that applying a temporal model for forecasting severe depressive states outperformed nontemporal models. Their study relied on a large-scale longitudinal data set of self-reported information about mood, activity, and sleep of 2382 self-declared depressed people over 22 months.

Cho et al [26] conducted a prospective observational cohort study to evaluate the mood of 55 patients with major depressive disorder and bipolar disorder types 1 and 2. They collected light exposure data passively via mobile phones of patients and self-reported daily mood scores. Using activity trackers, they registered activity, sleep, and heart rate data. This information was then processed into 130 features based on circadian rhythms, and mood prediction was performed using the RF method. Their approach generally showed good sensitivity and specificity for mood state and episode prediction.

Taylor et al [27] focused on building personalized models for forecasting the next day's mood (good or bad), health (fair or poor), and stress intensity (low or high). The multitask learning-based approach used data about the physiology and behavior of 206 undergraduate students and the weather of the current day, collected for 30 days. Their results showed that tomorrow's well-being could be predicted with 78% to 82% accuracy using a personalized model based on the present day's data. Busk et al [28] proposed a hierarchical Bayesian approach for forecasting mood for up to 7 days from smartphone self-assessments of 84 patients diagnosed with bipolar disorder. Their best performing model used a history of 4 days of self-assessment, indicating that short-term historical mood is a significant predictor.

Another recent observational study by Darvariu et al [29] combined user-reported emotional information, passive sensing data, and visual context information from individuals' surroundings in the form of images to develop deep learning techniques for emotional state inference. Their findings showed context-dependent associations between self-reported emotional states and the objects surrounding the individuals.

These studies provide insight into the potential of using mobile sensor data to infer individuals' mental well-being. However, none of these studies reported working with a data set consisting of observations from a nonexperimental setting or dealing with large amounts of missing data. Moreover, in most of these studies, the problem they are trying to solve is a 2-class classification problem. Here, the problem is approached from a more refined perspective (ie, predicting emotional state in terms of both valence and arousal dimensions).

Objectives

This study focuses on applying machine learning algorithms to predict mood states based on passively sensed behavioral patterns. Specifically, we aim to assess which behavioral features provide the most important information about daily emotional valence. The study was conducted by using data collected via a clinically validated eHealth platform (eB2 MindCare) [30,31]. This app is designed to run unobtrusively in the background of an individual's smartphone. It automatically and continuously gathers information about behavior, collected via both the individual's smartphone and wearable devices. It also provides an electronic diary type interface for users to register information about their emotions and other important events.

Methods

Participants

The data for this experiment were collected via eB2 MindCare [32] in collaboration with public mental health hospitals Hospital Universitario Fundación Jiménez Díaz and Hospital Universitario Rey Juan Carlos, Madrid, from clinical outpatients [30,33,34] and nonpathological volunteers from Universidad Carlos III de Madrid and Universidad Católica de Valencia. Patients were invited to participate in the data collection process by their clinicians. The research followed the code of ethics defined in the Declaration of Helsinki by the World Medical Association.

Patient Inclusion and Exclusion Criteria

Patients were included in the study if they were at least 18 years old clinical outpatients diagnosed by specialists at the institutions mentioned above with mental disorders or were attending therapy groups (such as support groups for cyberbullying and relaxation) at these institutes. They had to own a smartphone running on Android or iOS operating systems, which they connected to a Wi-Fi network at least more than once per week. Only patients who provided written informed consent for the eB2 study were included. None of the patients were paid for participating in the study.

Data

The eB2 MindCare app collects data from different sources (the mobile phone's sensors, Google Fit, and wearables such as Fitbit and Garmin) at different time intervals. After installing the app, the users are taken through an onboarding phase, in which they are asked to give permission for specific data collection streams, depending on the operating system. In addition to passively collected data, users can record information about their experiences, quality of sleep, and emotional state during the day. The app offers the following emotion options to choose from: angry, disgusted, scared, sad, overwhelmed, tired, grief, neutral, relaxed, motivated, happy, and delighted. Within a day, patients may register their emotions multiple times.

Daily summary values of 6 passively collected observations were considered: step count, distance traveled, hours of sleep, hours of phone use, time spent at home, and the number of locations visited. An additional binary variable was included,

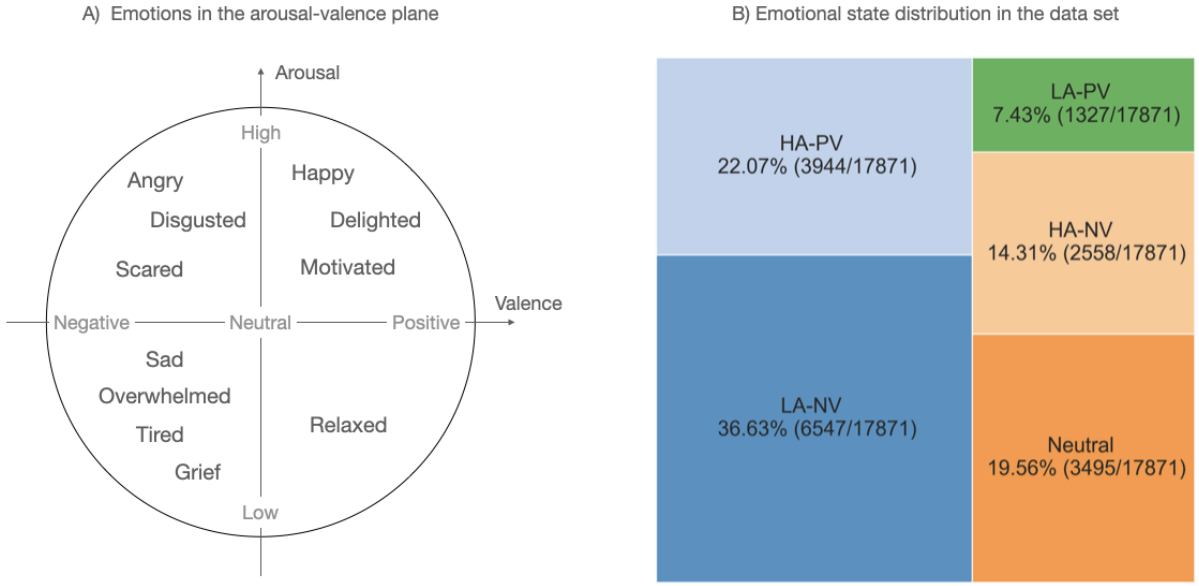
indicating whether the patient practiced sports during the day. The step count is recorded every 5 minutes, and the daily summary value corresponded to the sum of the registered entries. App use information was gained similarly. Distance information is gathered every minute, whereas location data are gathered at 5-minute intervals. Locations are obfuscated with an offset and randomly rotated to protect users' data. From these sources, the daily travel distance and the number of visited locations were computed. Time spent at home was computed using clustering based on the most common user locations throughout the day. There is a hierarchical set up for hours of sleep for the credibility of different sources; if data are manually introduced by the user or calculated by the phone but confirmed by the user, that value is first considered. Otherwise, the following ordering holds: sleep data by iOS, sleep data by Garmin, sleep data by Fitbit, sleep data calculated from light, app use and steps data, and sleep data calculated by the phone. The devices register sport-related activities on change, and the daily summary encompasses the total number of times each action was performed.

A subset of 943 users (patients and nonpathological subjects) was selected with at least 30 days of passively sensed data in the eB2 database between January 2019 and March 2020. The number of recorded days per patient varied from 30 to 487 with a mean of 190 (SD 122). Demographic information was available only for 871 users. All the users were Spaniards. Of these, 63.5% (553/871) were female and 25.1% (219/871) were male, and gender information was not available for the remaining 11.4% (99/871). All age groups were adequately represented in the data set, with a mean age of 41 years (range 18-77 years) computed at the beginning of the measurement period. The patient population came from 2 main categories: 61.3% (534/871) were outpatients from external psychiatric consultancy and 22.1% (192/871) were suicidal high-risk outpatients. The remaining 16.6% (145/871) were nonpathological users. Note that neither demographic nor diagnostic information was used in the rest of the study.

A well-known framework for dealing with emotional experience characterizes emotions in a 2-dimensional space defined by Russel [15,16]. The arousal and valence are combined, with valences ranging from highly negative to highly positive and arousal ranging from low to high. Daily emotional valence and arousal metrics were determined using raw emotion data entered by patients. Valence was then computed as the sign of the difference between positive and negative emotion counts, whereas arousal was determined based on the categories in the study by Scherer [35].

The left subfigure in Figure 1 shows the projection of emotions to the arousal-valence plane. The emotions listed on the graph are those that patients can register via the eB2 app. As the right subfigure in Figure 1 shows, there is a significant imbalance between the different emotional labels. The majority correspond to negative emotional valence (9105 entries), followed by positive emotions (5271 entries) and only 3495 neutral entries in the entire data set. Moreover, as emotions are self-reported, with users not being prompted in any way to fill in this information, these entries are scarce compared with passively sensed behavioral data.

Figure 1. Projection of emotions into the arousal-valence plane and their distribution in the data set. HA-NV: high arousal-negative valence; HA-PV: high arousal-positive valence; LA-NV: low arousal-negative valence; LA-PV: low arousal-positive valence.



As data have been collected from several sources and received in different formats, the raw daily summary data have many anomalies and unwanted information, and hence, noise. The presence of noise in the data can degrade the performance of machine learning methods. Therefore, it is important to preprocess the data before using it as an input to any machine learning algorithm. The first step of preprocessing was removing any negative values, followed by thresholding the time-related variables to 24 hours, the step count to 30,000 steps per day, and the distance to 500 km. Data were then standardized over all patient sequences, making each input feature 0 mean (SD 1).

Moreover, the data set contained a large percentage of missing observations. Figure S1 in Multimedia Appendix 1 shows the missing pattern in the entire data set. Approximately 84% (150,615/179,740) of the observations were partial, a bit over 5% (9399/179,740) were complete, and the remaining 10% (19,726/179,740) were entirely missing. Slightly less than 10% (17,871/179,740) of the observations were labeled by an emotion entry. A total of 271 patient sequences were observed for all 7 summary variables. Close to half of them did not have information about the time spent at home and the number of locations visited. The app use information was also completely missing for 226 patients. In addition, 114 patients had more than 30 consecutive days of completely missed observations (range 31-372).

Probabilistic Generative Models for Dealing With Missing Data

Imputing missing data using statistical measures such as the mean, median, or even interpolation fails when the percentage of missing data is very high. These approaches can reduce variability in the data set and introduce bias. However, probabilistic generative models can learn the underlying distributions in a data set by adjusting the model parameters to best account for the data in the sense of maximizing the evidence, even in the presence of missing data. Mixture models (MMs) [36] and hidden Markov models (HMMs) [37] are frequently used types of such models.

MMs comprise a finite or infinite number of components, possibly different distributional types, that can describe different data features. The data can then be modeled in terms of a mixture of several components, where each component has a simple parametric form (such as a Gaussian). The model is formulated in terms of latent variables, which represent the component each data point was sampled from and learned from the observed features, referred to as observables by adjusting the model parameters, which define the observable emission probabilities, such that the MM best accounts for the data in the sense of maximizing the evidence.

HMMs are temporal MMs that are commonly used for time-series analysis. These are generative models characterized by a set of observable sequences. The discrete states of the HMM are assumed to have been generated by a first-order Markov chain process, and each observation depends only on the paired state. An HMM comprises an initial state probability distribution, a state transition probability distribution, and a symbol emission probability distribution. Both MMs and HMMs were trained using the expectation-maximization algorithm.

In this study, the observed data were heterogeneous. Practice sport and emotional state are categorical, and the rest of the variables are assumed to be real-valued. Both MMs and HMMs can deal with missing data, without requiring imputation before training, via marginalization. For the Gaussian parameters, the diagonal covariance matrices were considered. Furthermore, both generative models were trained in a semisupervised manner for emotional valence and arousal-valence discrete observations. Namely, the different emotional states' emission probabilities were fixed for some of the components, whereas others were adjusted during training, such as the other model parameters. For instance, in a 5-component MM with binary label emissions, the emission probability for label 0 of the 3 components can be set to 1, forcing the components always to emit label 0. In contrast, the other 2 components can always be forced to emit label 1.

Emotion Prediction Models

A series of experiments were conducted for emotional status prediction accuracy using both nontemporal and temporal machine learning models. The underlying motivation was to analyze whether there were long-term dependencies in the data concerning patients' daily emotional states.

Probabilistic generative models (MM and HMM) were used to perform the imputation. Note that only the input features were imputed, and the emotion labels were not. When using MMs, first, for each observation, the posterior distribution needs to be inferred to find which component the observation is most likely to belong to; then, the missing attributes are imputed by a sample generated from that component. Information about the emotional state belonging to the current observation was not included in the posterior computation (otherwise, the model would overfit). When using HMMs, all observation sequences were first decoded using the Viterbi algorithm on the trained HMM. This method finds the most likely sequence of components that could have resulted in the given observation sequence. Once the state sequence was determined, the missing data were imputed by the samples generated from the corresponding states for each time step. The state posterior probabilities were computed by applying the forward algorithm [37], leaving out the current emotional observation.

For nontemporal machine learning methods, LR, support vector classifier, random forest classifier (RFC), and multilayer perceptron (MLP) were considered. These models allow comparisons with previous emotional state studies [16-19,26]. A grid search was performed for each case for hyperparameter tuning.

RNNs [38] have recurring inputs to the hidden layer; this allows them to remember input states from previous time steps, which can carry important information for future time-step predictions. There are 3 common types of RNNs: vanilla RNN, LSTM [39], and gated recurrent units (GRUs) [40]. Vanilla RNNs have short-term memory. If the observation sequence is rather long, these models have difficulty remembering relevant information from earlier time steps. LSTM and GRU cells, which contained gates that regulate the information flow, were designed to solve this problem.

In this experiment, RNNs of each of the 3 types were tested. A single layer with 64 hidden units was used, whose output was connected to a dense layer. Finally, the softmax activation function provides the predictions. The model was trained using the Adam method and the negative-log-likelihood loss for 50 epochs, using early stopping. One-layer RNNs with vanilla RNN, LSTM, and GRU cells were trained using 64 hidden units for each case. More complex models, such as dilated RNN, multilayer RNN, and temporal convolutional networks, have also been tried. However, they did not improve performance,

proving that simpler RNNs could explain the data's temporal correlations.

Personalized Models

To improve the above models, hierarchical Bayesian regression models were proposed to account for individual differences and predict the emotional state of patients. The proposed model allows intercepts to vary across patients, according to a random effect, while having a fixed slope for the predictor (ie, all patients will have the same slope). In our model for individual j, observation i, target variable y_{ji}, and input features x_{ji}:

where the random intercept effect is drawn from the population distribution:

The population mean and SD are independent normal and half-normal priors. By setting a separate bias term for each patient, rather than fitting separate regression models for each patient, multilevel modeling shares strength among patients, allowing for a more reasonable inference in patients with little data. The models were trained with Stein Variational Gradient Descent [41,42] for 50 epochs using the Adam optimizer.

Evaluation and Interpretability

Accuracy, area under the receiver operating characteristics curve (AUC-ROC), and area under the precision-recall curve (AUC-PRC) were used as the evaluation metrics. AUC-ROC is commonly used for both balanced and imbalanced classification problems because it is not biased toward the majority or minority class. However, AUC-PRC scores provide more insight into the minority class when the problem is very skewed. As the AUC-ROC and AUC-PRC scores are computed for binary classification problems, in the case of multiclass targets, different types of averaging can be performed on the data. The reported results were microaveraged, meaning that the metrics are global, computed by counting the total number of true positives, false negatives, and false positives.

On the basis of several model interpretability methods, Lundberg and Lee [43] defined the Shapley additive explanations (SHAP) value, a modality to explain any machine learning model's output. The SHAP values can provide global interpretability to the machine learning models by showing how much each feature contributes, positively or negatively, to the target variable. This approach was used in this study to analyze the feature importance for the models. Moreover, this method can be applied to analyze the decisions for individual predictions, which provides better insights into the relationships between passively collected mobile data and self-reported emotions.

Experiments

For MM and HMM training, only those patient sequences with at least partial observations for each of the 7 features and emotions were used. Moreover, the maximum sequence length was limited to 365 days, and sequences that had more than 30 days of consecutive missing data for all variables were discarded. After this elimination process, 233 sequences were used to train both the MMs and HMMs with different numbers of states. These patient sequences were excluded from both the training and test sets of the later models.

For the global models, the data set containing the remaining 710 patient sequences was divided into training and test sets using 80% of the sequences for training and 20% for testing. These data sets were kept independent. The train-test split cannot be done for the personalized models by randomly selecting a given percentage of the patients for training while leaving the others for testing, but all 710 patients must be included. Therefore, the patient sequences themselves were split into training and test sections. The first 80% of the labeled observations, in chronological order, were used for training, and the remaining samples were used for testing.

As the LR, support vector classifier, RFC, and MLP cannot directly exploit time-series data, we created the following 2 cases as inputs for these models. First, the input-output pairs consisted of 1 day of labeled observation. Second, 3 days and a week before the entered emotion was considered and concatenated into a single feature vector. In the case of the temporal models, training was performed with 30-day, 3-month, and 6-month long sequences.

Before creating the above feature vectors, the missing data in each patient sequence were imputed by the MM or HMM samples. For models trained with mini-batch stochastic gradient descent, every data point is imputed every time it enters the optimizer. The sequences were decoded multiple times, and missing data were imputed by samples generated from the corresponding state.

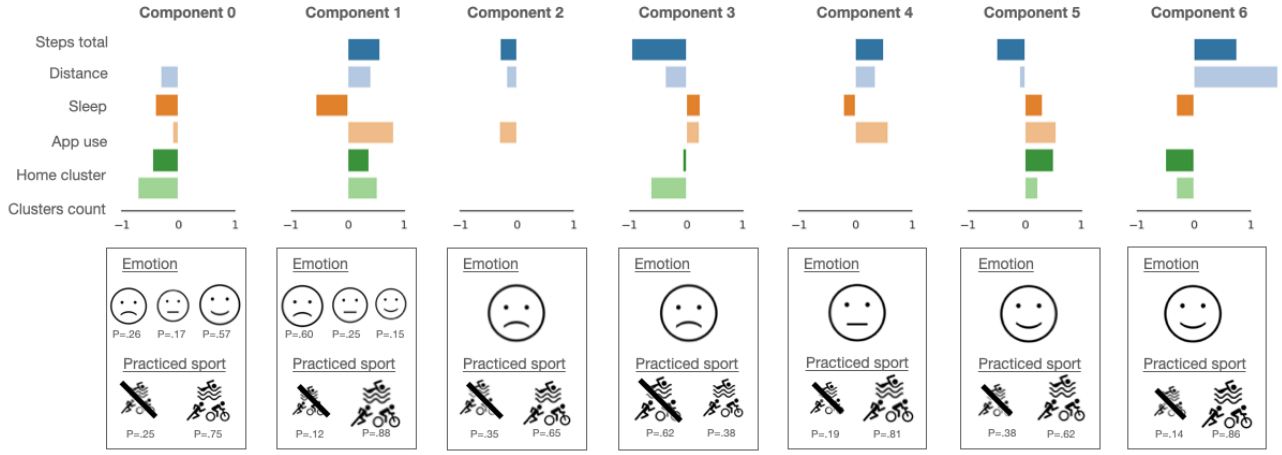
We designed two types of experiments. The first type is limited to the projection of the recorded emotions to a single axis of the arousal-valence plane. The second set of experiments considered 2-dimensional projections. A total of 3 different settings were analyzed for the classifiers' input features, as follows: using the imputed raw data, using the MM or HMM posterior probabilities instead of the raw input features, and using the raw inputs concatenated with the MM or HMM posterior probabilities.

Results

Generative Models

After experimenting with several hidden state setups, 7 hidden components captured the data's underlying patterns well, leading to the best results when a classifier was applied to the data later to predict emotions and provide interpretable states. In this case, the emission probabilities of the five states were fixed so that two pairs of states always emitted negative and positive emotions, and one always emitted a neutral emotion. The different components turned out to be specialized, as they captured contrasting behaviors (Figure 2).

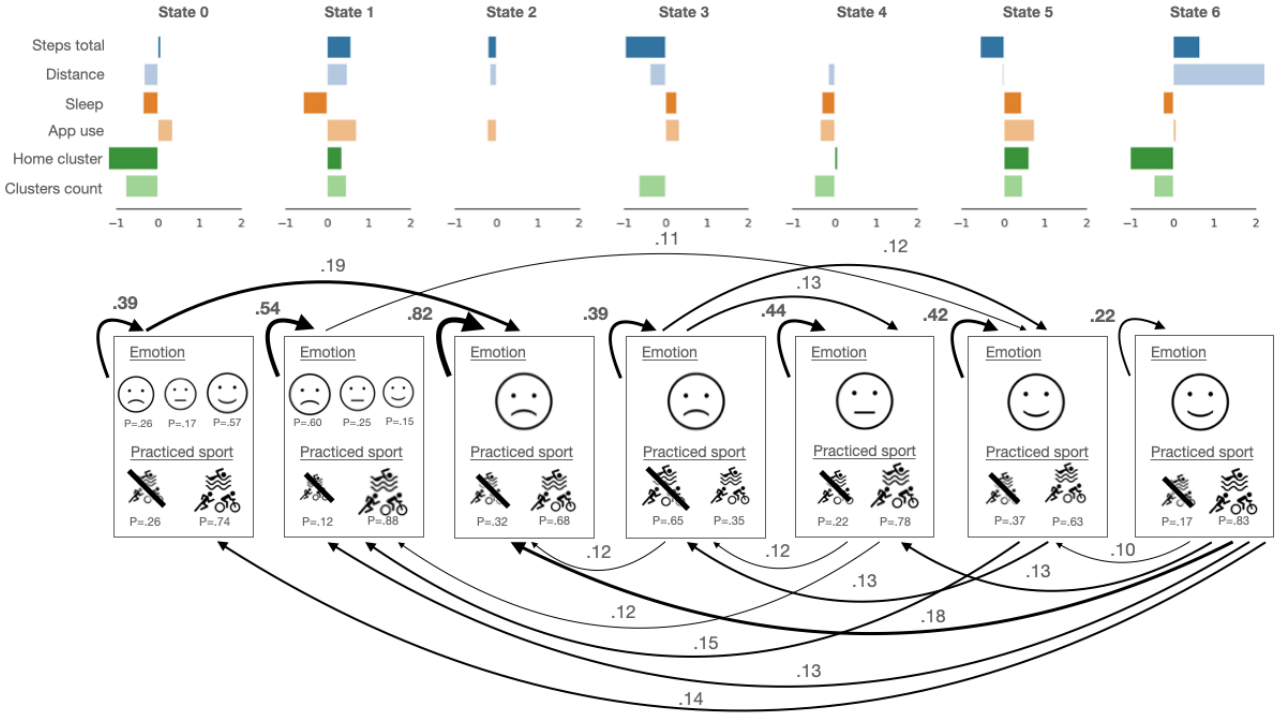
Figure 2. The 7-component mixture model structure was used for emotional valence modeling with each Gaussian mean in each component and indicating discrete emission probabilities. The size of the icons indicates the magnitude of the discrete emission probabilities (emotion and sport). In terms of features, "steps total" refers to step count, "distance" refers to the distance traveled, "sleep" refers to the hours of sleep, "app use" refers to the hours spent using different apps, "home cluster" refers to the time spent at home, "clusters count" refers to the number of visited locations, and "practiced sport" is an indicator of whether the patient practiced any sports. Of note, the negative mean values were a result of the normalization of the features.



Focusing on the three components that mainly emit negative emotional valence (components 1, 2, and 3), it can be seen that the corresponding modeled behaviors are contrasting. Component 1 represents days when the patients are quite active, visit multiple locations, spend a significant amount of time using their phones, and sleep very few hours. Component 2 is characterized by fewer steps and low app use. Component 3, however, captures days with low activity and mostly spent at home. The corresponding sport-related discrete emissions show that the patients practice some sport (>15 minutes of walking, biking, running, other, or a combination of those) in components 1 and 2, but less likely in component 3. Components 0, 5, and 6 correspond to positive emotional valence. They also seem to capture significantly different behavioral patterns. In component 0, the patients seemed to sleep less and did not spend much time at home; component 5 captured days with more time spent at home and excessive phone use. Component 6 captures the days of travel. Finally, the component capturing neutral emotions indicates days with medium activity and more app use.

Including the temporal properties of HMMs, the trained generative model with 7 hidden states and the same fixed emotional state emissions led to very similar interpretable outcomes as the MM (Figure 3). The temporal characteristics were not very strong. States 2 (with fixed negative emotional valence emission) and 1 (with mainly negative emotional valence emission) had the highest self-transition probabilities. If the self-transition probabilities are large, it indicates a stable state. States 0, 3, 4, and 5 have somewhat large self-transition probabilities, which suggests that days with positive and negative but also neutral emotions following each other are common in the patient population.

Figure 3. The 7-component hidden Markov model structure was used for emotional valence modeling with each Gaussian mean in each component and indicating discrete emission probabilities. The size of the icons indicates the magnitude of the discrete emission probabilities (emotion and sport). Only the transitions with a higher than 0.1 probability are shown in the graph. In terms of features, "steps total" refers to step count, "distance" refers to the distance traveled, "sleep" refers to the hours of sleep, "app use" refers to the hours spent using different apps, "home cluster" refers to the time spent at home, "clusters count" refers to the number of visited locations, and "practiced sport" is an indicator of whether the patient practiced any sports. Of note, the negative mean values were a result of the normalization of the features.



In the arousal-valence case, the 7-state generative models had 1 state assigned to all the emotional state emissions, and the other 2 were trained with the rest of the parameters. Similarly, as before, the states appear to capture specific behaviors, such as days of medium activity but mostly spent at home, more active days, days with more travel, and so on (Figures S2 and S3 in Multimedia Appendix 1 provide the sketches of the 7-component MM and HMM, respectively).

Predicting Emotional Valence

Figure S1 in Multimedia Appendix 2 compares the accuracy and the microaverage AUC-ROC and AUC-PRC scores for the trained classifiers in the 3 experimental set ups, as described in the *Experiments* section. Most classifiers achieved significantly higher performance than random guessing (AUC-ROC=0.5). As the figure shows, the models perform the worst on the raw data. Using the HMM or MM posteriors as input features or combining the raw data with the posteriors increases the performance. Table 1 compares the best performing models when using the MM and HMM posteriors. The difference in the results obtained with the MM posteriors and HMM posteriors is minimal. This indicates that the temporal dimension is not very relevant to the problem at hand; hence, a simpler generative model is sufficient for the problem.

Table 1. Performance comparison of the best performing models using mixture model and hidden Markov model posteriors as classifier input features.

Model and classifier input features	Accuracy (%)	Area under the receiver operating characteristics curve	Area under the precision-recall curve
Multilayer perceptron using 7 days of observations as input features			
Mixture model posteriors	65	0.81	0.70
Hidden Markov model posteriors	64	0.80	0.69

The best performing model was the MLP with the posteriors of 7 days of observations as input features. Concatenating the posterior probabilities for 3 days or 7 days of observations significantly improves the performance; however, training RNNs with longer observation sequences leads to decreased performance. This suggests no substantial seasonality or long-term trend of the self-reported emotions; thus, time-series models are not needed for the emotional state prediction task.

Generally, the most misclassified emotional state is the neutral state (refer to Table S1 in Multimedia Appendix 2 for confusion matrices). In most cases, it is confused with a negative emotional state and reasonably often with a positive one. There is some confusion between positive and negative emotional states, but somewhat fewer for negative emotions. This suggests that the models are more sensitive to detecting negative emotions, which can be desirable; for example, if the app's goal is to detect periods when the patient is feeling down.

Predicting Emotional Arousal Valence

In the second experiment, the target variables were the emotion projections into the 2-dimensional arousal-valence space, based on the categories in the study by Scherer [35]. Hence, the problem becomes a 5-class classification task. Here, we aimed to test the possibility of predicting daily emotions on a finer scale than the 2-class valence analysis presented above.

The best performance for the emotional arousal-valence prediction, with 48% accuracy (compared with the baseline of 20%), 0.77 AUC-ROC, and 0.50 AUC-PRC, was obtained by the RFC with 7 days of data concatenated with the MM posteriors. The GRU network trained on 30-day sequences reached results closest to those from the temporal models: 42% accuracy, 0.69 AUC-ROC, and 0.36 AUC-PRC. In this setting, the added MM posteriors' effect was more significant than the emotional valence prediction case. Using the posteriors as input features led to a 23% performance increase in some of the models. Table S2 in Multimedia Appendix 2 provides a detailed performance comparison of the models.

Predicting more refined emotional states is a difficult task, as not only are there more classes to distinguish, but the class imbalance is also more accentuated. The trained models became somewhat biased toward the majority classes, resulting in the wrong classification of the minority classes (high arousal-positive valence and low arousal-positive valence). Generally, when the predictor variable is well separable, and there are no overlaps between the different classes, this separation can compensate for the imbalance; however, in this data set, that is not the case. Standard techniques to combat the imbalance problem, such as upsampling of minority classes, downsampling of majority classes, and one-versus-rest training, were applied; however, these only led to a slight improvement. Therefore, these results have not been reported.

Personalized Models

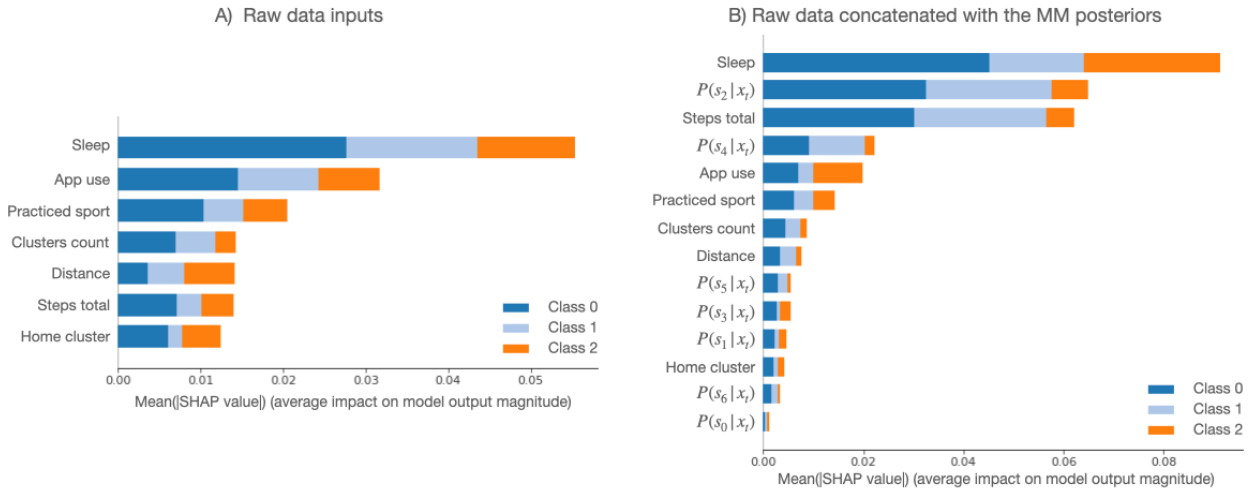
The previously presented models try to explain the variability of the observations by considering the patient population. As shown before, these models do not provide enough diversity when the classifier takes as input 1-day worth of data. Personalized models can provide a more scalable and accurate way to achieve better representations for individual patients.

The posterior probabilities obtained from the MM components were used as input features for the personalized models because they proved to improve the prediction outputs of earlier experiments. In the global models presented previously, features representing 1 day of data led to insufficient classifier accuracy, especially in the LR models, which only reached a maximum of 43% for the 3-class problem and 16% for the 5-class problem. The proposed hierarchical Bayesian LR method led to a significant increase in performance, reaching 64% accuracy, 0.81 AUC-ROC, and 0.70 AUC-PRC for the 3-class problem and 52% accuracy, 0.82 AUC-ROC, and 0.55 AUC-PRC for the 5-class problem. This demonstrates that accounting for individual differences through a simple hierarchical model can substantially improve emotional state prediction performance without relying on previous days of data.

Feature Importance Analysis

Figure 4 provides an overview of which features are most important for the emotional valence MLP models using the raw data and using the raw data and MM posteriors as input features. To obtain an overview of which features are most important for the models, the mean SHAP values (*Evaluation* and *Interpretability* section) of every feature for every sample were computed. The plot below sorts features by the mean absolute value of the SHAP value magnitudes over all samples.

Figure 4. Summary plot of feature importance for the multilayer perceptron models for emotional valence prediction, showing raw data and raw data concatenated with mixture model posteriors. In terms of features, "steps total" refers to step count, "distance" refers to the distance traveled, "sleep" refers to the hours of sleep, "app use" refers to the hours spent using different apps, "home cluster" refers to the time spent at home, "clusters count" refers to the number of visited locations, "practiced sport" is an indicator of whether the patient practiced any sports, and "$P(s_i \mid x_t)$" refers to the posterior probability in component *i*. The following class labels were used: 0=negative; 1=neutral; and 2=positive emotional valence. MM: mixture model; SHAP: Shapley additive explanations.

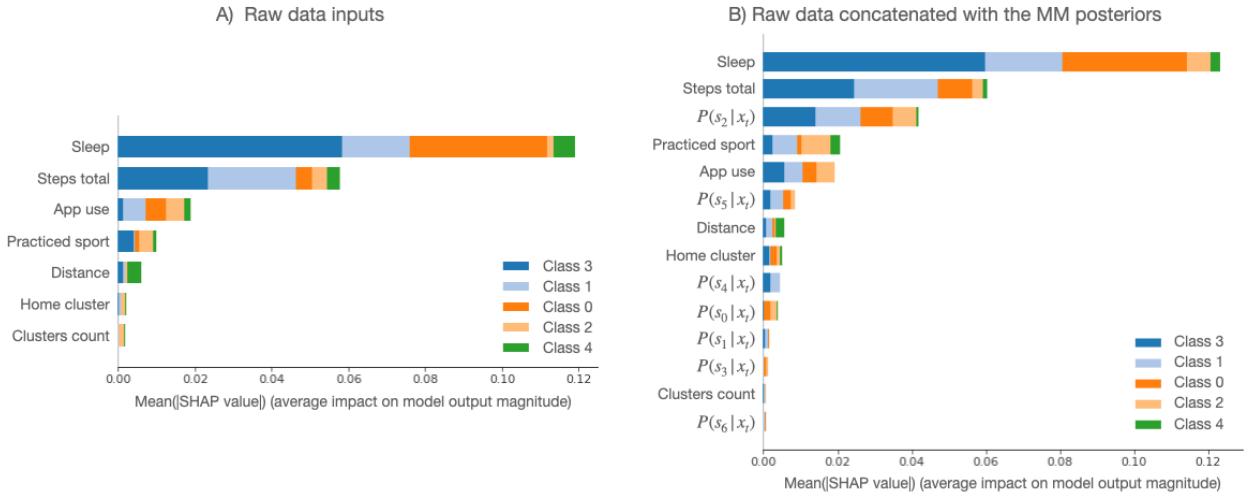


The hours of sleep and the time spent using their phone (app use) influenced all classes' outcomes the most. The other features have an almost similar influence on the positive and negative classes. The negative output (class 0) is also strongly

influenced by the step count, sport indicator, and time spent at home. If the posterior probabilities are used in combination with the raw features as inputs to the model, some outweigh the raw features in the decision-making process. For instance, the MLP relies heavily on the hours of sleep, the posterior probability of state 2, and the step count. The other classes seem to be more involved, requiring several posterior probabilities and raw values to form the prediction. The importance of posterior probabilities underlines the robust feature extraction provided by MM.

Similarly, the arousal-valence classifiers can be analyzed. In the raw data case (Figure 5), although the model emphasizes the hours of sleep and the step count, the other parameters become slightly less important. In the second case (Figure 5), some of the posterior probabilities seem to weigh more in the decision-making process than the raw features, as in the first experiment.

Figure 5. Summary plot of feature importance for the random forest models for emotional arousal-valence prediction, showing raw data and raw data concatenated with mixture model posteriors. In terms of features, "steps total" refers to step count, "distance" refers to the distance traveled, "sleep" refers to the hours of sleep, "app use" refers to the hours spent using different apps, "home cluster" refers to the time spent at home, "clusters count" refers to the number of visited locations, "practiced sport" is an indicator of whether the patient practiced any sports, and "$P(s_i \mid x_t)$" represents the posterior probability in component *i*. The following class labels were used: 0=neutral; 1=high arousal-positive valence; 2=high arousal-negative valence; 3=low arousal-negative valence; and 4=low arousal-positive valence. MM: mixture model; SHAP: Shapley additive explanations.



Discussion

Principal Findings

A variety of different machine learning methods were used to analyze passively sensed behavioral data from 6 sources (step count, distance traveled, hours of sleep, hours of phone use, time spent at home, number of locations visited, and a binary variable indicating whether the patient practiced sports during the day). These models were used to predict self-reported emotional state (valence or combination of valence and arousal) in a large, heterogeneous sample of treatment-seeking patients with clinically significant levels of psychological and/or emotional symptoms. Preliminary inspection of this data set revealed that the data exhibited significant missingness (approximately 84% [150615/179740] of the observations were partial). This represents real-world clinical data sets, which usually contain many missing samples and are sparsely labeled. The fact that this kind of data are both noisy and often nonrandomly missing means that the development of robust imputation techniques is a nontrivial problem. However, the development of such methods is vital if this type of information is used to support clinical decision making.

We addressed this problem by training generative models to handle missing data. These models were then used for data imputation and latent state (feature) extraction for emotional state prediction. Notably, predictive models performed significantly better when MM or HMM posterior probabilities were included alongside the raw behavioral input features. This suggests that the latent representation of the passively sensed behavioral variables discovered by the probabilistic generative models contains information relevant to daily emotional experience fluctuations. However, using HMMs over MMs did not improve the classification performance, which implies that there are no strong temporal correlations in the daily observations that can be captured by an HMM. Furthermore, in both experiments, the nonlinear models outperformed the other static models. The use of RNNs did not improve daily emotion predictions, suggesting that long-term behavior does not significantly influence patients' everyday emotional states.

When using raw data alone as input features, the hours of sleep had the most substantial influence on the emotional state predictions. The importance of activity-related features varied between the 2 experimental set ups. When posterior state probabilities were included in the model, some proved to be more important than the raw features. This indicates that the MM provided excellent feature representation and filtering of the observed behavioral signals. Interestingly, an inspection of the confusion matrices for the best performing models revealed that, for the valence prediction analysis, models were more sensitive to the detection of negative, compared with positive or neutral emotional states. This is a useful feature as this is the domain of emotional experience most likely to be relevant for clinicians or self-monitoring of trends in overall mental health.

Finally, we proposed a hierarchical Bayesian regression with varying intercepts and a common slope to personalize the models. This approach performs personalized predictions while accounting for population-level characteristics. The personalized models using 1-day long feature vectors achieved similar performance to the nonlinear variants using 3-day long feature vectors. Moreover, they performed significantly better than global linear LR models. Personalized models outperforming the generalized models are intuitively reasonable as mood is very personal, and its perceptions among individuals differ.

Limitations

This study has some limitations. As previously mentioned, the data analyzed in this study contained a large percentage of missing observations; approximately 84% (150,615/179,740) of the observations were partial, only a bit over 5% (9399/179,740) were complete, and the remaining 10% (19,726/179,740) were entirely missing. Some of the patient sequences had large chunks of consecutively missing observations, which could be because of sensor or software errors or the patients not using their devices for an extensive amount of time. Moreover, information about emotional states was sporadically reported. Therefore, only 10% (17,871/179,740) of the behavioral data were labeled with respect to the outcome of interest.

Recording emotions is a subjective process, and regular reflection of the emotional state may influence how one answers. The majority of the registered emotions were negatively valenced, meaning that the prediction models were somewhat biased toward negative emotional states. As a result, the models were most sensitive in the negative domain, and the overall prediction accuracies were not high in some cases. In addition, mood variability, another important point in psychiatric disorders, was not analyzed in this study. However, it will be important to explore in the future to better differentiate whether it is a pathological mood state or a mood within the normal range.

Conclusions

This study is an initial step toward developing more robust and informed models for predicting emotional states from passively sensed data. It presents a sound basis for further exploration by proposing a solution to missing and sparsely labeled data, allowing the future focus to be directed toward developing more advanced models.

Future plans include examining other deep learning models to improve prediction accuracy and analyzing effects at a more refined time scale. Another intriguing question is to consider the effect of seasonality (weekdays and weekends, seasonal variation) on patients' emotional states. Moreover, the possibilities of specialized models for different patient groups or individual patients will be further investigated.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under the Marie Sklodowska-Curie grant agreement number 813533. This work was partly supported by the Spanish government (Ministerio de Ciencia e Innovación) under grants TEC2017-92552-EXP and RTI2018-099655-B-100; the Comunidad de Madrid under grants IND2017/TIC-7618, IND2018/TIC-9649, IND2020/TIC-17372, and Y2018/TCS-4705; the BBVA Foundation under the Domain Alignment and Data Wrangling with Deep Generative Models **(**Deep-DARWiN) project; and the European Union (European Regional Development Fund and the European Research Council) through the European Union's Horizon 2020 Research and Innovation Program under grant 714161. The authors thank Enrique Baca-Garcia for providing demographic and clinical data and assisting in interpreting and summarizing the data.

Authors' Contributions

ES, with the supervision and guidance of AA and PMO, designed and conducted the study. AN and MMP-R provided expert advice on clinical and patient-related matters. All authors contributed to the writing and editing of the manuscript.

Conflicts of Interest

AA is co-founder of Evidence-Based Behavior (eB2). MMP-R has received research grant funding from Neurocrine Biosciences (Inc), Millennium Pharmaceuticals, Takeda, Merck, and AI Cure. She is an Advisory Board member for Neurocrine Biosciences, Inc, and a consultant on an American Foundation for Suicide Prevention (AFSP) grant (LSRG-1-005-16, PI: Baca-Garcia).

Multimedia Appendix 1
Supplementary figures of the data missingness pattern and the mixture model and hidden Markov model structures for the emotional arousal-valence case.
[PDF File (Adobe PDF File), 316 KB - mhealth_v9i3e24465_app1.pdf]

Multimedia Appendix 2
An overview of the accuracy, the microaverage area under the receiver operating characteristics curve and area under the precision-recall curve scores, and confusion matrices for each of the models in the 3 experimental setups, as described in the

Experiments section, for both the 3-class and 5-class experiments using the mixture model for missing data imputation and feature extraction.
[PDF File (Adobe PDF File), 157 KB - mhealth_v9i3e24465_app2.pdf]

References

1. Ortiz A, Grof P. Electronic monitoring of self-reported mood: the return of the subjective? Int J Bipolar Disord 2016 Dec 29;4(1):28 [FREE Full text] [doi: 10.1186/s40345-016-0069-x] [Medline: 27900735]
2. Griffeth BT, Brooks WB, Foster A. A psychiatric-specific entrustable professional activity for the evaluation of prospective psychiatric residents: towards a national standard. MedEdPORTAL 2017 May 23;13(1) [FREE Full text] [doi: 10.15766/mep_2374-8265.10584] [Medline: 30800786]
3. Panlilio LV, Stull SW, Kowalczyk WJ, Phillips KA, Schroeder JR, Bertz JW, et al. Stress, craving and mood as predictors of early dropout from opioid agonist therapy. Drug Alcohol Depend 2019 Sep 01;202:200-208 [FREE Full text] [doi: 10.1016/j.drugalcdep.2019.05.026] [Medline: 31357121]
4. Anestis MD, Selby EA, Crosby RD, Wonderlich SA, Engel SG, Joiner TE. A comparison of retrospective self-report versus ecological momentary assessment measures of affective lability in the examination of its relationship with bulimic symptomatology. Behav Res Ther 2010 Jul;48(7):607-613 [FREE Full text] [doi: 10.1016/j.brat.2010.03.012] [Medline: 20392437]
5. Gershon A, Kaufmann CN, Torous J, Depp C, Ketter TA. Electronic Ecological Momentary Assessment (EMA) in youth with bipolar disorder: demographic and clinical predictors of electronic EMA adherence. J Psychiatr Res 2019 Sep;116:14-18. [doi: 10.1016/j.jpsychires.2019.05.026] [Medline: 31176107]
6. Burke LE, Shiffman S, Music E, Styn MA, Kriska A, Smailagic A, et al. Ecological momentary assessment in behavioral research: addressing technological and human participant challenges. J Med Internet Res 2017 Mar 15;19(3):e77 [FREE Full text] [doi: 10.2196/jmir.7138] [Medline: 28298264]
7. LiKamWa R, Liu Y, Lane ND, Zhong L. MoodScope: building a mood sensor from smartphone usage patterns. In: Proceedings of the 11th annual international conference on Mobile systems, applications,services (MobiSys '13). 2013 Presented at: 11th annual international conference on Mobile systems, applications,services (MobiSys '13); June, 2013; Taipei Taiwan p. 389-402. [doi: 10.1145/2462456.2464449]
8. Mehrotra A, Tsapeli F, Hendley R, Musolesi M. MyTraces: investigating correlation and causation between users? Emotional states and mobile phone interaction. Proc ACM Interact Mob Wearable Ubiquitous Technol 2017 Sep 11;1(3):1-21. [doi: 10.1145/3130948]
9. Alvarez-Lozano J, Osmani V, Mayora O, Frost M, Bardram J, Faurholt-Jepsen M, et al. Tell me your apps and I will tell you your mood: correlation of apps usage with bipolar disorder state. In: Proceedings of the 7th International Conference on PErvasive Technologies Related to Assistive Environments - PETRA '14. 2014 Presented at: 7th International Conference on PErvasive Technologies Related to Assistive Environments - PETRA '14; May, 2014; Rhodes Greece p. 1-7. [doi: 10.1145/2674396.2674408]
10. Canzian L, Musolesi M. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15. Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp ?15 New York, New York, USA: ACM Press; 2015 Presented at: 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15; September, 2015; Osaka Japan p. 1293-1304. [doi: 10.1145/2750858.2805845]
11. Mehrotra A, Musolesi M. Using autoencoders to automatically extract mobility features for predicting depressive states. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Sep 18;2(3):1-20. [doi: 10.1145/3264937]
12. Choudhury MD, De CM, Counts S, Horvitz E. Predicting postpartum changes in emotion and behavior via social media. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI ?13; April, 2013; Paris France p. 3267-3276. [doi: 10.1145/2470654.2466447]
13. Lu H, Frauendorfer D, Rabbi M, Mast M, Chittaranjan G, Campbell A, et al. StressSense: detecting stress in unconstrained acoustic environments using smartphones. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12. 2012 Presented at: ACM Conference on Ubiquitous Computing - UbiComp '12; September, 2012; Pittsburgh, Pennsylvania p. 351-360. [doi: 10.1145/2370216.2370270]
14. Ma Y, Xu B, Bai Y, Sun G, Zhu R. Daily mood assessment based on mobile phone sensing. In: Proceedings of the Ninth International Conference on Wearable and Implantable Body Sensor Networks. 2012 Presented at: Ninth International Conference on Wearable and Implantable Body Sensor Networks; May 9-12, 2012; London, UK. [doi: 10.1109/bsn.2012.3]
15. Russell JA. Core affect and the psychological construction of emotion. Psychol Rev 2003 Jan;110(1):145-172. [doi: 10.1037/0033-295x.110.1.145] [Medline: 12529060]
16. Russell JA. A circumplex model of affect. J Pers Soc Psychol 1980;39(6):1161-1178. [doi: 10.1037/h0077714]

17. Jaques N, Taylor S, Azaria A, Ghandeharioun A, Sano A, Picard R. Predicting students' happiness from physiology, phone, mobility, and behavioral data. Int Conf Affect Comput Intell Interact Workshops 2015 Sep;2015:222-228 [FREE Full text] [doi: 10.1109/ACII.2015.7344575] [Medline: 28515966]
18. Sano A, Taylor S, McHill AW, Phillips AJ, Barger LK, Klerman E, et al. Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: observational study. J Med Internet Res 2018 Jun 08;20(6):e210 [FREE Full text] [doi: 10.2196/jmir.9410] [Medline: 29884610]
19. Umematsu T, Sano A, Taylor S, Picard R. Improving students' daily life stress forecasting using LSTM neural networks. In: Proceedings of the IEEE EMBS International Conference on Biomedical & Health Informatics (BHI). 2019 Presented at: IEEE EMBS International Conference on Biomedical & Health Informatics (BHI); May 19-22, 2019; Chicago, IL, USA. [doi: 10.1109/bhi.2019.8834624]
20. Morshed MB, Saha K, Li R, D'Mello SK, De Choudhury M, Abowd GD, et al. Prediction of mood instability with passive sensing. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Sep 09;3(3):1-21. [doi: 10.1145/3351233]
21. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, et al. StudentLife: using smartphones to assess mental health and academic performance of college students. In: Mobile Health. Switzerland: Springer; 2017:7-33.
22. Mattingly SM, Gregg JM, Audia P, Bayraktaroglu AE, Campbell AT, Chawla NV, et al. The Tesserae Project: Large-Scale, Longitudinal, In Situ, Multimodal Sensing of Information Workers. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI Conference on Human Factors in Computing Systems; May 2019; Glasgow, Scotland, UK p. 1-8. [doi: 10.1145/3290607.3299041]
23. Pollak JP, Adams P, Gay G. PAM: a photographic affect meter for frequent, in situ measurement of affect. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011 Presented at: Proceedings of the annual conference on Human factors in computing systems - CHI ?11; May, 2011; Vancouver BC, Canada p. 725-734. [doi: 10.1145/1978942.1979047]
24. Servia-Rodríguez S, Rachuri KK, Mascolo C, Rentfrow PJ, Lathia N, Sandstrom GM. 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: Proceedings of the 26th International Conference on World Wide Web; April, 2017; Perth, Australia. [doi: 10.1145/3038912.3052618]
25. Suhara Y, Xu Y. Deepmood: forecasting depressed mood based on self-reported histories via recurrent neural networks. In: Proceedings of the 26th International Conference on World Wide Web. 2017 Presented at: 26th International World Wide Web Conference; April, 2017; Perth, Australia. [doi: 10.1145/3038912.3052676]
26. Cho CH, Lee T, Kim MG, In HP, Kim L, Lee HJ. Mood prediction of patients with mood disorders by machine learning using passive digital phenotypes based on the circadian rhythm: prospective observational cohort study. J Med Internet Res 2019 Apr 17;21(4) [FREE Full text] [doi: 10.2196/11029] [Medline: 30994461]
27. Taylor SA, Jaques N, Nosakhare E, Sano A, Picard R. Personalized multitask learning for predicting tomorrow's mood, stress, and health. IEEE Trans Affect Comput 2020;11(2):200-213. [doi: 10.1109/TAFFC.2017.2784832] [Medline: 32489521]
28. Busk J, Faurholt-Jepsen M, Frost M, Bardram JE, Kessing VL, Winther O. Forecasting mood in bipolar disorder from smartphone self-assessments: hierarchical bayesian approach. JMIR Mhealth Uhealth 2020 Apr 01;8(4) [FREE Full text] [doi: 10.2196/15028] [Medline: 32234702]
29. Darvariu VA, Convertino L, Mehrotra A, Musolesi M. Quantifying the relationships between everyday objects and emotional states through deep learning based image analysis using smartphones. Proc ACM Interact Mob Wearable Ubiquitous Technol 2020 Mar 18;4(1):1-21 [FREE Full text] [doi: 10.1145/3380997]
30. Berrouiguet S, Ramírez D, Barrigón ML, Moreno-Muñoz P, Carmona Camacho R, Baca-García E, et al. Combining continuous smartphone native sensors data capture and unsupervised data mining techniques for behavioral changes detection: a case series of the evidence-based behavior (eB2) study. JMIR Mhealth Uhealth 2018 Dec 10;6(12):e197 [FREE Full text] [doi: 10.2196/mhealth.9472] [Medline: 30530465]
31. Carretero P, Campana-Montes JJ, Artes-Rodriguez A. Ecological momentary assessment for monitoring risk of suicide behavior. Curr Top Behav Neurosci 2020;46:229-245. [doi: 10.1007/7854_2020_170] [Medline: 32797403]
32. Evidence-based behavior. URL: https://eb2.tech/?lang=en [accessed 2021-02-23]
33. Barrigón ML, Berrouiguet S, Carballo JJ, Bonal-Giménez C, Fernández-Navarro P, Pfang B, MEmind study group. User profiles of an electronic mental health tool for ecological momentary assessment: MEmind. Int J Methods Psychiatr Res 2017 Mar;26(1) [FREE Full text] [doi: 10.1002/mpr.1554] [Medline: 28276176]
34. Porras-Segovia A, Molina-Madueño RM, Berrouiguet S, López-Castroman J, Barrigón ML, Pérez-Rodríguez MS, et al. Smartphone-based ecological momentary assessment (EMA) in psychiatric patients and student controls: a real-world feasibility study. J Affect Disord 2020 Sep 01;274:733-741. [doi: 10.1016/j.jad.2020.05.067] [Medline: 32664009]
35. Scherer KR. What are emotions? And how can they be measured? Soc Sci Info 2016 Jun 29;44(4):695-729. [doi: 10.1177/0539018405058216]
36. Bishop CM. Pattern recognition and machine learning. New York: Springer-Verlag; 2006.
37. Rabiner LR. A tutorial on hidden markov models and selected applications in speech recognition. In: Readings in Speech Recognition. Amsterdam: Elsevier; 1990:267-296.

38. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature 1986 Oct 9;323(6088):533-536. [doi: 10.1038/323533a0]
39. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997 Nov 15;9(8):1735-1780. [doi: 10.1162/neco.1997.9.8.1735] [Medline: 9377276]
40. Cho K, van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, et al. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014 Presented at: Conference on Empirical Methods in Natural Language Processing (EMNLP); October 2014; Doha, Qatar p. 1724-1734. [doi: 10.3115/v1/d14-1179]
41. Liu Q, Wang D. Stein variational gradient descent: a general purpose Bayesian inference algorithm. In: Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS'16). 2016 Presented at: 30th International Conference on Neural Information Processing Systems (NIPS'16); 2016; Red Hook, NY, USA.
42. Liu Y, Ramachandran P, Liu Q, Peng J. Stein variational policy gradient. 2017. URL: http://arxiv.org/abs/1704.02399 [accessed 2021-02-23]
43. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). 2017 Presented at: the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc.. Red Hook, NY, USA.?4777; 2017; Red Hook, NY, USA p. 4768-4777.

Abbreviations

AUC-PRC: area under the precision-recall curve
AUC-ROC: area under the receiver operating characteristics curve
Deep-DARWiN: Domain Alignment and Data Wrangling with Deep Generative Models
EMA: ecological momentary assessment
GRU: gated recurrent unit
HMM: hidden Markov model
LR: logistic regression
LSTM: long short-term memory
mHealth: mobile health
MLP: multilayer perceptron
MM: mixture model
RF: random forest
RFC: random forest classifier
RNN: recurrent neural network
SHAP: Shapley additive explanations
SVM: support vector machine

Edited by M Focsa; submitted 23.09.20; peer-reviewed by CH Cho, S Steingrimsson, J Seitz; comments to author 21.11.20; revised version received 29.01.21; accepted 15.02.21; published 22.03.21.

Please cite as:
Sükei E, Norbury A, Perez-Rodriguez MM, Olmos PM, Artés A
Predicting Emotional States Using Behavioral Markers Derived From Passively Sensed Data: Data-Driven Machine Learning Approach
JMIR Mhealth Uhealth 2021;9(3):e24465
URL: https://mhealth.jmir.org/2021/3/e24465
doi:10.2196/24465
PMID:33749612

Original Paper

Measuring Criterion Validity of Microinteraction Ecological Momentary Assessment (Micro-EMA): Exploratory Pilot Study With Physical Activity Measurement

Aditya Ponnada[1], BDes; Binod Thapa-Chhetry[1], MSc; Justin Manjourides[2], PhD; Stephen Intille[1], PhD

[1]Khoury College of Computer Sciences, Bouve College of Health Sciences, Northeastern University, Boston, MA, United States

[2]Bouve College of Health Sciences, Northeastern University, Boston, MA, United States

Corresponding Author:
Aditya Ponnada, BDes
Khoury College of Computer Sciences
Bouve College of Health Sciences
Northeastern University
360 Huntington Avenue
Boston, MA,
United States
Phone: 1 617 306 1610
Email: ponnada.a@northeastern.edu

Abstract

Background: Ecological momentary assessment (EMA) is an in situ method of gathering self-report on behaviors using mobile devices. In typical phone-based EMAs, participants are prompted repeatedly with multiple-choice questions, often causing participation burden. Alternatively, microinteraction EMA (micro-EMA or µEMA) is a type of EMA where all the self-report prompts are single-question surveys that can be answered using a 1-tap glanceable microinteraction conveniently on a smartwatch. Prior work suggests that µEMA may permit a substantially higher prompting rate than EMA, yielding higher response rates and lower participation burden. This is achieved by ensuring µEMA prompt questions are quick and cognitively simple to answer. However, the validity of participant responses from µEMA self-report has not yet been formally assessed.

Objective: In this pilot study, we explored the criterion validity of µEMA self-report on a smartwatch, using physical activity (PA) assessment as an example behavior of interest.

Methods: A total of 17 participants answered 72 µEMA prompts each day for 1 week using a custom-built µEMA smartwatch app. At each prompt, they self-reported whether they were doing sedentary, light/standing, moderate/walking, or vigorous activities by tapping on the smartwatch screen. Responses were compared with a research-grade activity monitor worn on the dominant ankle simultaneously (and continuously) measuring PA.

Results: Participants had an 87.01% (5226/6006) µEMA completion rate and a 74.00% (5226/7062) compliance rate taking an average of only 5.4 (SD 1.5) seconds to answer a prompt. When comparing µEMA responses with the activity monitor, we observed significantly higher (P<.001) momentary PA levels on the activity monitor when participants self-reported engaging in moderate+vigorous activities compared with sedentary or light/standing activities. The same comparison did not yield any significant differences in momentary PA levels as recorded by the activity monitor when the µEMA responses were randomly generated (ie, simulating careless taps on the smartwatch).

Conclusions: For PA measurement, high-frequency µEMA self-report could be used to capture information that appears consistent with that of a research-grade continuous sensor for sedentary, light, and moderate+vigorous activity, suggesting criterion validity. The preliminary results show that participants were not carelessly answering µEMA prompts by randomly tapping on the smartwatch but were reporting their true behavior at that moment. However, more research is needed to examine the criterion validity of µEMA when measuring vigorous activities.

(JMIR Mhealth Uhealth 2021;9(3):e23391) doi:10.2196/23391

KEYWORDS

ecological momentary assessment (EMA); experience sampling; physical activity; smartwatch; microinteractions; criterion validity; activity monitor; µEMA

Introduction

Ecological momentary assessment (EMA), also known as the experience sampling method, is used to measure behaviors of people in natural settings [1]. In a typical EMA study, a user's phone is prompted multiple times a day (often 6+ times) with a set of multiple-choice questions measuring behaviors of interest [2,3]. The repeated EMA prompts, which typically ask about momentary behaviors or states, not only reduce recall biases present in retrospective surveys [3,4] but also capture temporal changes in health behaviors unique to each individual [5]. Because of these benefits, EMA is commonly used to measure behaviors in intensive longitudinal studies [6].

The drawback of EMA is participation burden [7-9]. Participants are first interrupted with a beep and/or vibration. They must then find the phone, unlock the device, and respond to a set of complex multiple-choice questions. This repeated effort, which can take tens of seconds for even the shortest surveys and several minutes for many common surveys, can be burdensome, negatively impacting study compliance [7,9,10]. Microinteraction EMA (µEMA or micro-EMA) is a type of EMA that may, for some behaviors (eg, chronic pain or fatigue), enable high frequency self-report data collection with low study burden [11]. In µEMA, rather than using complex multiquestion surveys, each prompt contains only a single question that can be answered with a glanceable microinteraction [12], typically just a tap on a smartwatch. Prior studies have shown that despite approximately 8 times more interruption than EMA, µEMA had a significantly higher response rate and lower perceived burden because all interactions are limited to microinteractions [11,13]. Thus, there is preliminary evidence that µEMA may enable gathering high-frequency self-report with manageable burden, a complementary approach to EMA. Recently, µEMA has been used to gather data on stress [14], hyperarousal [15], and perceived comfort [16], and it has also been used with small pervasive displays [17].

Prior work on µEMA, however, has assumed validity of µEMA responses and not demonstrated it. Because µEMA is designed to gather small amounts of information with each prompt (but with higher frequency), the prompts are both limited to a single question and made cognitively simple to answer with a quick microinteraction (taking only 3-5 seconds). To achieve cognitive simplicity and fit questions on a smartwatch so they can be answered in a single tap without scrolling requires a limited answer set. This calls into question whether µEMA responses could capture behavior similarly to a gold-standard instrument (ie, criterion validity [18]), and validating such a high-frequency self-report requires an instrument that can measure the same behavior continuously in free living, such as a wearable sensor. In some domains where µEMA may be especially useful (eg, chronic pain), such sensors do not yet exist. The purpose of this pilot study is to explore the criterion validity of µEMA self-report, and thus, we used the example of physical activity (PA) measurement, because PA can be estimated continuously using research-grade activity monitors [19].

Methods

In this pilot study, we compared µEMA self-report on a smartwatch with acceleration data collected using a wearable activity monitor on the dominant ankle to assess criterion validity, such as whether participants are answering the µEMA questions meaningfully in a way that changes as PA changes.

µEMA App

We implemented a µEMA app on an Android Wear OS 2.0 (Figure 1) to measure PA. PA was chosen because (1) it can be estimated continuously using a passive, easy-to-wear sensor (eg, an accelerometer on the ankle) and (2) PA can change frequently within a day, making it suitable for testing a high-frequency µEMA self-report system. Participants were presented with 4 activity intensity options with each µEMA prompt: sedentary, light/standing, moderate/walking, and vigorous.

Figure 1. (Left) µEMA interface on a smartwatch with four activity intensity options. (Right) Undo screen to change response, available for 10 s, with a countdown timer.

The µEMA app prompted 6 times an hour between 8 AM to 8 PM (72 expected prompts per day) using vibration on the smartwatch. The question displayed at the start of the vibration consisted of the 4 activity intensity categories, and participants selected the activity intensity they were engaged in at that moment. If the participant did not respond to the prompt within 2 minutes by tapping on a category, the prompt disappeared from the screen and the app recorded a missed response. When a response was selected, the watch displayed an Undo? screen (Figure 1). If participants tapped on the Undo? button, they were returned to the µEMA question to change their responses, otherwise the Undo? screen disappeared after 10 seconds. Data from the watch were sent to the participant's smartphone once per hour. Participants interacted only with the watch, and the phone collected, encrypted, and transferred data from the watch to a remote server.

Study Design

We conducted a week-long, within-subject pilot study (approved by Northeastern University's institutional review board; project number 14-10-01) to compare µEMA PA responses with a wearable sensor.

Participant Recruitment

Participants were eligible if they owned a compatible Android smartphone with version 4.3+, were aged 18 to 55 years, were a student or staff at our university (to ensure we could safely recover loaned smartwatches and activity monitors at the end of the study), and were willing to wear a sensor and the smartwatch for 1 week. The study was advertised using flyers posted on the university campus and also by sending electronic notices to common university announcement portals. Of the 35 people who responded to study advertisements, 20 were eligible to participate based on screening via phone call. Among those, 3 participants dropped out early. One had a wake-period outside of µEMA prompting hours, another had a job that physically made it difficult to answer prompts on the watch, and the third had a malfunctioning phone. This left 17 active participants in the pilot study (11 males and 6 females; aged 19 to 34 years). None of the participants were affiliated to our research group.

Measurement Tasks

Participants were asked to complete 2 tasks simultaneously between 8 AM to 8 PM every day for one week: answer µEMA prompts on the loaned smartwatch (model Urbane, LG Electronics) and wear an activity monitor on their dominant ankle. Participants were asked to ignore µEMA prompts in unsafe conditions (eg, driving) and charge the watch nightly so that it could be worn next morning.

We used GT9X monitors (35×35 mm, 14 g; ActiGraph LLC) to measure acceleration continuously at 80 Hz [20]. Participants wore the sensor on the dominant ankle above the medial malleolus using an elastic band (Figure 2). The ankle location was chosen because ankle acceleration can reliably capture ambulation activities, more so than the wrist or hip [21,22]. The sensor collected raw acceleration passively; other than wearing it, the participants did not interact with, or charge, this device.

Figure 2. GT9X activity monitor worn on the dominant ankle.

Procedures

Day 0: Researchers met with participants, obtained informed consent, and loaned the activity monitor and smartwatch with the µEMA app. Research staff then presented participants with some examples of the types of different activities that would fall into each of the 4 target activity categories (Multimedia Appendix 1). Days 1-7: Participants wore the activity monitor on the dominant ankle and answered 6 µEMA prompts per hour between 8 AM to 8 PM on the smartwatch for 1 week. Day 8: Researchers recovered the monitor. Participants were not compensated financially to ensure that we measure criterion validity without any external motivation, but participants could use the smartwatch for 4 more days as their personal device for fun if they desired. All participants agreed to use the device for 4 more days.

Results

We computed participants' (1) compliance rate, the percentage of µEMA prompts answered out of all the scheduled prompts (ie, including when the watch was off); (2) completion rate, the percentage of µEMA prompts answered out of delivered prompts (ie, excluding when the watch was off); and (3) response time, the time taken to answer a prompt, measured from the start of the prompt vibration.

Two participants had low compliance (Figure 3). From the debriefing, we learned that these participants did not charge the smartwatch regularly, receiving fewer scheduled prompts. Their compliance fell below the 1.5 interquartile range (<40%); therefore, they were considered outliers and were excluded from the main analysis of data from the remaining 15 participants (Table 1) [23]. Implications of dropping the outliers is discussed later in our results.

Figure 3. µEMA compliance and completion rates.

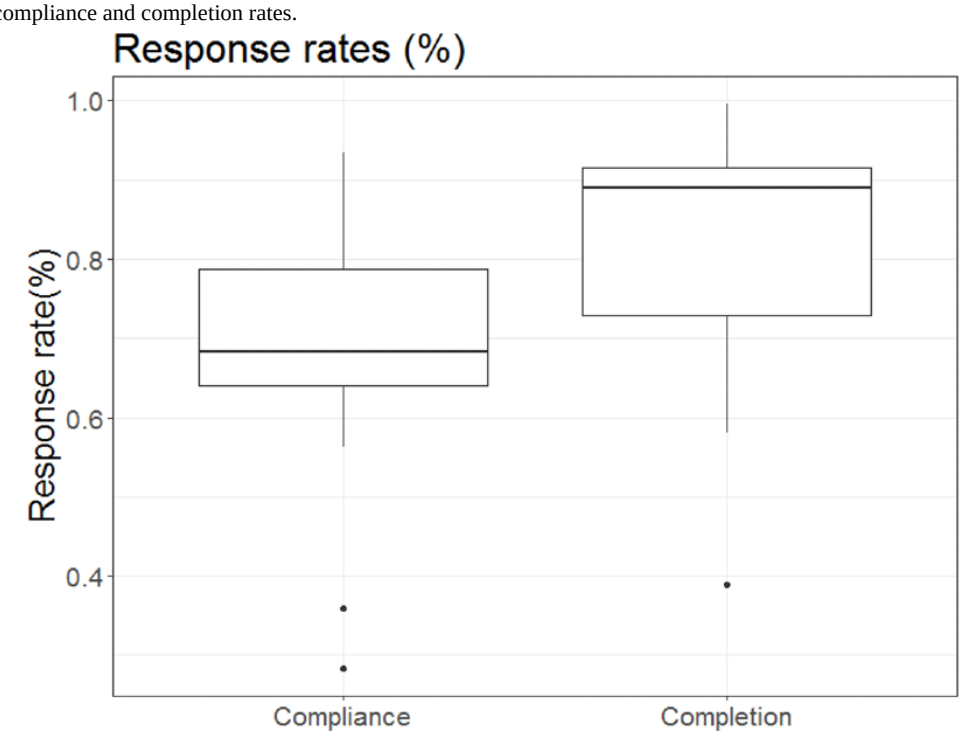


Table 1. Response behavior of pilot study participants.

Characteristics	With outliers (n=17)	Without outliers (n=15)
Expected prompts, n	7591	7062
Delivered prompts, n	6387	6006
Answered prompts, n	5238	5226
Compliance rate (%)	69.00	74.00
Completion rate (%)	82.01	87.01
Response time (s), mean (SD)	5.5 (1.6)	5.4 (1.5)

Data Preparation

Computing Activity Counts

ActiGraph activity counts are widely used motion summary metrics computed from raw acceleration for a specified epoch [24]. Activity counts have been used in prior work to compare EMA responses and accelerometer data [25,26]. We first computed activity counts for 1 second epochs of the raw data from the ankle-worn activity monitor. Using this, we calculated the total activity counts 60 seconds prior to the µEMA prompt as the PA level measured using the activity monitor.

Removing Sensor Nonwear Data

We removed the instances of raw data when the participants were not wearing the activity monitor. Following Choi et al [27], 90+ minutes of continuous zero-valued activity counts computed for the 1 second epochs were considered sensor nonwear times. All µEMA responses recorded during these nonwear times were dropped. This eliminated only 1.3% (70/5226) of the total responses from 15 participants leaving 5156 valid responses with sensor wear. Of these, 25% (18/70) of the responses were from just 1 participant.

µEMA Response Distribution

We received more sedentary responses (3619/5156, 69.99%) than light/standing (978/5156, 18.97%) and moderate/walking (544/5156, 10.56%) from µEMA, which is consistent with general physical (in)activity trends [28,29]. However, we received few vigorous responses (15/5156, 0.29%). Thus, we combined the moderate/walking and vigorous categories into a single category of moderate+vigorous. Hence, we compare the 3 PA intensities (sedentary, light/standing, and moderate+vigorous) from µEMA with activity count (60 seconds prior to the prompt) from the ankle-worn activity monitor (Table 2). These activity counts were within the ranges recorded previously in young adults for sedentary, light, and moderate+vigorous activities using ankle-worn accelerometers [30-33].

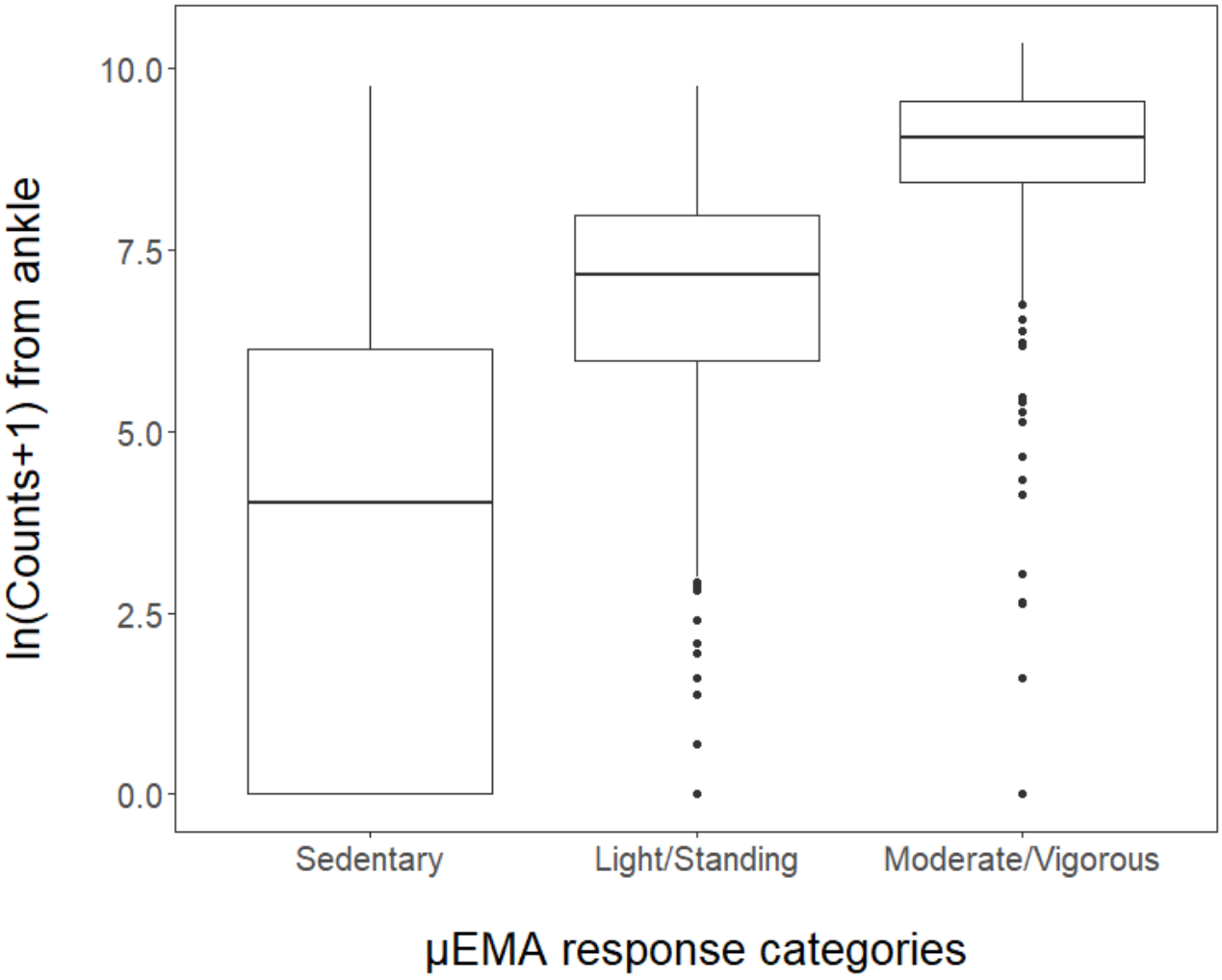
Activity counts computed 60 seconds before the prompt ranged from 0 (for sedentary) to >15K (for moderate/vigorous), and resulted in a right-skewed distribution. Thus, we log-transformed these activity counts into ln(Counts + 1), where 1 is the smallest nonzero count recorded in this pilot study [34]. Figure 4 presents the final distribution of these log-transformed counts corresponding to µEMA categories (sedentary, light/standing, and moderate+vigorous).

Table 2. ln(Counts + 1) measured on ankle for each µEMA category.

Category	Value
Sedentary	
Mean (SD)	3.56 (2.94)
Median (IQR[a])	4.01 (6.13)
Light/standing	
Mean (SD)	6.66 (2.10)
Median (IQR)	7.16 (2.01)
Moderate+vigorous	
Mean (SD)	8.72 (1.52)
Median (IQR)	9.04 (1.11)

[a]IQR: interquartile range.

Figure 4. µEMA responses versus ln(Counts + 1).



Criterion Validity of µEMA on Smartwatch

We applied a linear mixed-effects model with a random intercept (using the lme4 package [35]):

$$\ln(Counts_{ij} + 1) = {}_0 + {}_1 (EMA_{ij}) + u_i$$

Here, $Counts_{ij}$ is the activity count from the ankle-worn monitor measured for an individual *i* computed for 60 seconds before µEMA prompt *j*, β_0 is the fixed-effect intercept, μEMA_{ij} is the ordinal self-report (0 = sedentary, 1 = light/standing, and 2 = moderate/vigorous) on the smartwatch by participant *i* at prompt *j*, and u_i is the random-intercept for the participant *i*. Although we were interested in the fixed-effects part of the model, the random intercept is included to account for repeated measures within participants. The momentary PA levels measured on the activity monitor, $\ln(Counts_{ij} + 1)$, were significantly different (*P*<.001) for sedentary, light/standing, and moderate+vigorous activity categories captured using µEMA self-report (ie, μEMA_{ij}). The final model fit for this pilot study was:

$$\ln(Counts_{ij} + 1) = 6.28 + 3.57 (EMA_{ij}) + u_i$$

We then included data from the two outliers to explore the sensitivity of the model fit and the activity levels, $\ln(Counts_{ij} + 1)$ were significantly different (*P*<.001) for the 3 µEMA response categories (μEMA_{ij}). However, we observed that these participants showed continuous sensor nonwear for 2 to 3 days at once in addition to not charging the smartwatch regularly as required by the study protocol. In fact, removing the sensor nonwear for these participants eliminated 27.78% (30/108) and 43.04% (65/151) of the answered µEMA prompts, respectively. As a result, these participants were excluded from our final model fit. For the remaining 15 participants, the pair-wise, post hoc comparison of the 3 µEMA response categories with Tukey adjustment revealed that $\ln(Counts_{ij} + 1)$ for sedentary responses were significantly less than light/standing, which were significantly less than the moderate+vigorous (*P*<.001), the expected order of activity intensity (Table 3). In other words, when participants self-reported (using µEMA) being in moderate+vigorous instead of light/standing or sedentary activities, they were also more likely to have passively measured higher PA levels in those moments.

For exploratory purposes, we simulated random µEMA responses for each participant; instead of analyzing their actual responses, we compared the random data as if participants were randomly tapping on the watch screen only to dismiss the prompt with motion data recorded from the ankle. The model fit did not yield any significant differences in the counts for different µEMA categories. This further suggests that

participants were answering µEMA questions carefully, not randomly (or carelessly), despite the intensive sampling rate.

Table 3. Pairwise comparison of ln(Counts + 1) from activity monitor with µEMA responses.

µEMA[a] response (i)	µEMA response (j)	Mean difference (i–j)	SE	95% CI
Sedentary	Light/standing	–3.12	0.28	–3.34 to –2.90
Light/standing	Moderate+vigorous	–1.93	0.13	–2.24 to –1.16
Moderate+vigorous	Sedentary	5.05	0.11	4.78 to 5.32

[a]µEMA: microinteraction ecological momentary assessment.

Discussion

Principal Findings

In this pilot study, we explored criterion validity of µEMA on a smartwatch by measuring PA intensity of free-living individuals. For 1 week, 15 participants answered 6 µEMA prompts per hour (between 8 AM to 8 PM each day) reporting their PA intensity while simultaneously wearing an ankle-worn research-grade activity monitor. The activity monitor gathered raw accelerometer data continuously at 80 Hz. When comparing the µEMA self-report and activity monitor, we observed significantly higher momentary PA levels on the activity monitor when participants self-reported engaging in moderate+vigorous activities compared with when they reported sedentary or light/standing activities. Similarly, we observed significantly higher PA levels on the activity monitor when participants self-reported engaging in light/standing activities compared with when they reported being sedentary. Thus, we observed the expected order of intensity of these activity categories, suggesting criterion validity of µEMA self-report. Holistically, this result is aligned with the prior work where PA researchers have compared phone-based EMA with body-worn research-grade activity monitors measuring raw acceleration and found EMA to measure PA similar to the objective sensor [25,36,37]. However, the frequency of EMA self-report in these studies varied from once a day [25] to once an hour [37] to minimize interruption burden. In our pilot study, we extend these findings to an alternate self-report approach (µEMA) that allowed for 6 times more temporal density in measurement, yielding high response rate but without burdening participants as much as the phone based EMA.

Overall, this preliminary result shows that when measuring PA, participants' µEMA self-report (at a high temporal density) could capture the PA levels consistent with a continuous high-frequency sensor. It appears, for example, that participants answered µEMA prompts meaningfully—not just tapping on an answer to dismiss the prompt—and sustained this answering at a rate of 6 times per hour for 12 hours per day for an entire week (approximately 504 prompts per person). Our preliminary findings suggest that the µEMA prompts achieve the goal of being so easy to answer that they can be sustained, instead of being ignored or dismissed—all while recording the true behavior at that moment. Participants know that every µEMA prompt just requires a single, 1-tap response on the easily accessible smartwatch that can be completed in a microinteraction; this may contribute to high compliance and valid data entry. EMA protocols, alternatively, often require answering multiple, sometimes complex, questions that can be time consuming and feel burdensome. In fact, the effort needed to dismiss an µEMA prompt is roughly equivalent to the effort required to answer the prompt, thus encouraging survey completion.

Limitations

This exploratory pilot study provides preliminary findings on criterion validity of µEMA, but more research is needed. We merged the vigorous and moderate/walking activities into moderate+vigorous because we received only 15 responses of vigorous activities. One reason may be that most individuals engage in significantly more sedentary than vigorous activities [38]. Another may be that during the exit debriefing, 2 participants reported that when they engaged in vigorous activities like outdoor cycling, they could not respond to prompts within 2 minutes, thus resulting in fewer vigorous activity responses. This type of response behavior during vigorous activities has also been observed in a prior EMA versus activity monitor validation study where more missing responses were found during vigorous activity [37]. Nevertheless, this potential for a bias when reporting vigorous activity using µEMA should be explored in future work. If a reporting bias for vigorous activity is observed in future studies, one remedy could be to explore sensor-triggered µEMA, where the µEMA prompts might be presented based on real-time processing of PA and then delivered not during vigorous PA but rather right after it is confidently estimated to have been completed [39]. This also highlights the need to rethink question wording, where instead of asking about behavior in the moment (eg, Doing vigorous PA now?), µEMA will have to ask about the recent past (eg, Vigorous PA 2 min ago?) but without compromising the cognitive simplicity required for a microinteraction. Notably, EMA delivered via a smartphone would have likely been even more difficult to respond to during vigorous PA due to the difficulty of accessing and interacting with the phone device (versus the comparative simplicity of completing a microinteraction on the smartwatch) [40]. Future studies with larger sample sizes including individuals recruited specifically because they are known to regularly engage in vigorous activities could provide more insights on µEMA validity when measuring vigorous activity.

Being an exploratory pilot study, we had a small sample size. We were limited by available equipment, and recruiting was more challenging than in typical studies because we did not

offer financial compensation for study participation (contrary to most phone EMA studies); our intent was to measure µEMA compliance and data quality in a situation without any external monetary incentives because such incentives may not be viable for future longitudinal measurement or intervention studies that might use µEMA.

Because validity assessments are domain-dependent, our findings with PA do not necessarily generalize to other behaviors. PA was chosen in this work because we have research-grade activity monitors to compare µEMA responses against at a high temporal density, not necessarily because PA intensity is best measured with µEMA if passive sensors are available. However, in some domains where µEMA might be useful (eg, chronic pain), passive sensors that continuously monitor the behavior are not available yet, and methods such as direct or physiological observations that require laboratory conditions are not practical for multiday free-living studies. Nevertheless, validation studies generally rely on imperfect comparisons, and so confidence in validity of µEMA, just like EMA, will require multiple cross-domain and longitudinal experiments from different research teams, of which this may be the first of many.

Conclusion

The µEMA implemented on a smartwatch addresses the common Achilles' heel of traditional phone-based EMA, the participation burden of accessing, unlocking, and answering multiple multichoice questions on a smartphone. Despite significantly higher interruption rates than phone-based EMA, µEMA is able to yield significantly higher response rates with manageable participation burden. This is achieved by keeping the single questions in µEMA cognitively simple to answer allowing for high frequency of prompting (like a continuous sensor). However, this makes µEMA vulnerable to careless tapping on the smartwatch to dismiss the prompts, potentially compromising the validity of self-report responses. Thus, in this pilot study, we explored the criterion validity of µEMA self-report, comparing it with a continuous sensor to assess if participants submit their responses based on their true behavior in the moment. We used PA as an example domain because PA can be measured using a gold-standard sensor (research-grade accelerometers). We conducted a 1-week exploratory pilot study with 15 participants answering 72 µEMA prompts each day while measuring continuous PA using an ankle-worn accelerometer. We found that participants were able to correctly report their sedentary, light/standing, and moderate+vigorous activities. This highlights that participants were not carelessly tapping on the smartwatch only to dismiss the prompt but were providing accurate information about their behavior comparable to the continuous sensor data from the ankle suggesting criterion validity. However, more research is needed to explore criterion validity of µEMA in other behavioral domains of interest including vigorous PA measurement.

Acknowledgments

AP led development of the µEMA data collection app and experiment design and wrote the first draft of the manuscript. BTC assisted with raw accelerometer data processing from the activity monitors. JM provided guidance on statistical analysis and interpreting the results. SI was the principal investigator for the project responsible for securing research funding and working closely with AP on project design and manuscript development. We thank Krystal Huey for assisting in participant recruitment and data collection and our participants for their time. Research reported in this paper was supported, in part, by a Google Glass Award and by the National Heart, Lung, and Blood Institute of the National Institutes of Health under award number R21HL108018. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Onboarding instructions.
[DOCX File , 14 KB - mhealth_v9i3e23391_app1.docx]

References

1. Stone AA, Shiffman S. Ecological momentary assessment: measuring real world processes in behavioral medicine. Ann Behav Med 1994;16:199-202. [Medline: 28298264]
2. Smyth J, Smyth JM. Ecological Momentary Assessment research in behavioral medicine. J Happiness Studies 2003;4(1):35-52 [FREE Full text] [doi: 10.1023/a:1023657221954]
3. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
4. Shiffman S. Ecological momentary assessment (EMA) in studies of substance use. Psychol Assess 2009 Dec;21(4):486-497 [FREE Full text] [doi: 10.1037/a0017074] [Medline: 19947783]

5. Chandra S, Scharf D, Shiffman S. Within-day temporal patterns of smoking, withdrawal symptoms, and craving. Drug Alcohol Depend 2011 Sep 01;117(2-3):118-125 [FREE Full text] [doi: 10.1016/j.drugalcdep.2010.12.027] [Medline: 21324611]
6. van Berkel N, Ferreira D, Kostakos V. The experience sampling method on mobile devices. ACM Comput. Surv 2018 Jan 12;50(6):1-40 [FREE Full text] [doi: 10.1145/3123988]
7. Collins RL, Kashdan TB, Gollnisch G. The feasibility of using cellular phones to collect ecological momentary assessment data: application to alcohol consumption. Exp Clin Psychopharmacol 2003 Feb;11(1):73-78. [doi: 10.1037//1064-1297.11.1.73] [Medline: 12622345]
8. Courvoisier DS, Eid M, Lischetzke T. Compliance to a cell phone-based ecological momentary assessment study: the effect of time and personality characteristics. Psychol Assess 2012 Sep;24(3):713-720. [doi: 10.1037/a0026733] [Medline: 22250597]
9. Fuller-Tyszkiewicz M, Skouteris H, Richardson B, Blore J, Holmes M, Mills J. Does the burden of the experience sampling method undermine data quality in state body image research? Body Image 2013 Sep;10(4):607-613 [FREE Full text] [doi: 10.1016/j.bodyim.2013.06.003] [Medline: 23856302]
10. Ram N, Brinberg M, Pincus AL, Conroy DE. The questionable ecological validity of ecological momentary assessment: considerations for design and analysis. Res Hum Dev 2017;14(3):253-270 [FREE Full text] [doi: 10.1080/15427609.2017.1340052] [Medline: 30613195]
11. Intille S, Haynes C, Maniar D, Ponnada A, Manjourides J. µEMA: microinteraction-based ecological momentary assessment (EMA) using a smartwatch. Proc ACM Int Conf Ubiquitous Comput 2016 Sep;2016:1124-1128 [FREE Full text] [doi: 10.1145/2971648.2971717] [Medline: 30238088]
12. Ashbrook D. Enabling Mobile Microinteractions [Dissertation]. Atlanta: Georgia Institute of Technology; 2010.
13. Ponnada A, Haynes C, Maniar D, Manjourides J, Intille S. Microinteraction ecological momentary assessment response rates: effect of microinteractions or the smartwatch? Proc ACM Interact Mob Wearable Ubiquitous Technol 2017 Sep;1(3) [FREE Full text] [doi: 10.1145/3130957] [Medline: 30198012]
14. King ZD, Moskowitz J, Egilmez B, Zhang S, Zhang L, Bass M, et al. Micro-stress EMA: a passive sensing framework for detecting in-the-wild stress in pregnant mothers. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Sep;3(3) [FREE Full text] [doi: 10.1145/3351249] [Medline: 32432212]
15. Larsen J, Eskelund K, Christiansen T. Active self-tracking of subjective experience with a one-button wearable: a case study in military PTSD. 2017 Presented at: International Conference on Human Factors in Computing Systems; 2017; Denver URL: https://arxiv.org/abs/1703.03437
16. Jayathissa P, Quintana M, Abdelrahman M, Miller C. Humans-as-a-sensor for buildings—intensive longitudinal indoor comfort models. Buildings 2020 Oct 01;10(10):174 [FREE Full text] [doi: 10.3390/buildings10100174]
17. Paruthi G, Raj S, Baek S, Wang C, Huang C, Chang Y, et al. Heed: exploring the design of situated self-reporting devices. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Sep 18;2(3):1-21 [FREE Full text] [doi: 10.1145/3264942]
18. Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychol Bull 1955 Jul;52(4):281-302. [Medline: 13245896]
19. Meijer GA, Westerterp KR, Verhoeven FM, Koper HB, ten Hoor F. Methods to assess physical activity with special reference to motion sensors and accelerometers. IEEE Trans Biomed Eng 1991 Mar;38(3):221-229. [doi: 10.1109/10.133202] [Medline: 2066134]
20. Miller J. Accelerometer technologies, specfications, and limitations: a presentation by ActiGraph at ICAMPAM 2013. Actigraph, LLC. 2013 Jun 17. URL: https://actigraphcorp.com/wp-content/uploads/2019/12/ActiGraph-ICAMPAM-2013.pdf [accessed 2021-03-02]
21. LaMunion S, Bassett D, Toth L, Crouter S. The effect of body placement site on ActiGraph wGT3X-BT activity counts. Biomed. Phys. Eng. Express 2017 Jun 23;3(3):035026 [FREE Full text] [doi: 10.1088/2057-1976/aa777c]
22. Mannini A, Sabatini A, Intille S. Accelerometry-based recognition of the placement sites of a wearable sensor. Pervasive Mob Comput 2015 Aug 01;21:62-74 [FREE Full text] [doi: 10.1016/j.pmcj.2015.06.003] [Medline: 26213528]
23. Pagano M, Gauvreau K. Principles of Biostatistics. London: Chapman and Hall; Nov 2018.
24. What are counts? Actigraph. 2018 Nov 08. URL: https://actigraphcorp.force.com/support/s/article/What-are-counts [accessed 2021-03-02]
25. Knell G, Gabriel KP, Businelle MS, Shuval K, Wetter DW, Kendzor DE. Ecological momentary assessment of physical activity: validation study. J Med Internet Res 2017 Jul 18;19(7):e253 [FREE Full text] [doi: 10.2196/jmir.7602] [Medline: 28720556]
26. Bruening M, van Woerden I, Todd M, Brennhofer S, Laska MN, Dunton G. A mobile ecological momentary assessment tool (devilSPARC) for nutrition and physical activity behaviors in college students: a validation study. J Med Internet Res 2016 Jul 27;18(7):e209 [FREE Full text] [doi: 10.2196/jmir.5969] [Medline: 27465701]
27. Choi L, Ward SC, Schnelle JF, Buchowski MS. Assessment of wear/nonwear time classification algorithms for triaxial accelerometer. Med Sci Sports Exerc 2012 Oct;44(10):2009-2016 [FREE Full text] [doi: 10.1249/MSS.0b013e318258cb36] [Medline: 22525772]

28. Matthews CE, Chen KY, Freedson PS, Buchowski MS, Beech BM, Pate RR, et al. Amount of time spent in sedentary behaviors in the United States, 2003-2004. Am J Epidemiol 2008 Apr 1;167(7):875-881 [FREE Full text] [doi: 10.1093/aje/kwm390] [Medline: 18303006]
29. Sisson SB, Camhi SM, Church TS, Martin CK, Tudor-Locke C, Bouchard C, et al. Leisure time sedentary behavior, occupational/domestic physical activity, and metabolic syndrome in U.S. men and women. Metab Syndr Relat Disord 2009 Dec;7(6):529-536 [FREE Full text] [doi: 10.1089/met.2009.0023] [Medline: 19900152]
30. Ozemek C, Kirschner MM, Wilkerson BS, Byun W, Kaminsky LA. Intermonitor reliability of the GT3X+ accelerometer at hip, wrist and ankle sites during activities of daily living. Physiol Meas 2014 Feb;35(2):129-138. [doi: 10.1088/0967-3334/35/2/129] [Medline: 24399138]
31. Rawson ES, Walsh TM. Estimation of resistance exercise energy expenditure using accelerometry. Med Sci Sports Exerc 2010 Mar;42(3):622-628. [doi: 10.1249/MSS.0b013e3181b64ef3] [Medline: 19952824]
32. Stec MJ, Rawson ES. Estimation of resistance exercise energy expenditure using triaxial accelerometry. J Strength Cond Res 2012 May;26(5):1413-1422. [doi: 10.1519/JSC.0b013e318248d7b4] [Medline: 22222328]
33. Rhudy MB, Dreisbach SB, Moran MD, Ruggiero MJ, Veerabhadrappa P. Cut points of the Actigraph GT9X for moderate and vigorous intensity physical activity at four different wear locations. J Sports Sci 2020 Mar;38(5):503-510. [doi: 10.1080/02640414.2019.1707956] [Medline: 31865845]
34. McCune B, Urban DL. Data transformations. In: McCune B, Grace JB, Urban DL, editors. Analysis of Ecological Communities. Gleneden Beach: MjM Software Design; 2002.
35. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Soft 2015;67(1):48. [doi: 10.18637/jss.v067.i01]
36. Dunton GF, Whalen CK, Jamner LD, Henker B, Floro JN. Using ecologic momentary assessment to measure physical activity during adolescence. Am J Prev Med 2005 Nov;29(4):281-287. [doi: 10.1016/j.amepre.2005.07.020] [Medline: 16242590]
37. Dunton GF, Liao Y, Kawabata K, Intille S. Momentary assessment of adults' physical activity and sedentary behavior: feasibility and validity. Front Psychol 2012;3:260 [FREE Full text] [doi: 10.3389/fpsyg.2012.00260] [Medline: 22866046]
38. Tucker JM, Welk GJ, Beyler NK. Physical activity in U.S.: adults compliance with the Physical Activity Guidelines for Americans. Am J Prev Med 2011 Apr;40(4):454-461. [doi: 10.1016/j.amepre.2010.12.016] [Medline: 21406280]
39. Dunton GF, Dzubur E, Intille S. Feasibility and performance test of a real-time sensor-informed context-sensitive ecological momentary assessment to capture physical activity. J Med Internet Res 2016 Jun 01;18(6):e106 [FREE Full text] [doi: 10.2196/jmir.5398] [Medline: 27251313]
40. Ashbrook D, Clawson J, Lyons K, Starner T, Patel N. Quickdraw: the impact of mobility and on-body placement on device access time. 2008 Presented at: Proc SIGCHI Conf Human Factors Comput Syst; 2008; Florence URL: https://doi.org/10.1145/1357054.1357092 [doi: 10.1145/1357054.1357092]

Abbreviations

EMA: ecological momentary assessment
PA: physical activity
μEMA: microinteraction ecological momentary assessment

Edited by L Buis; submitted 11.08.20; peer-reviewed by M Newman, A Miller; comments to author 06.10.20; revised version received 01.12.20; accepted 22.02.21; published 10.03.21.

Please cite as:
Ponnada A, Thapa-Chhetry B, Manjourides J, Intille S
Measuring Criterion Validity of Microinteraction Ecological Momentary Assessment (Micro-EMA): Exploratory Pilot Study With Physical Activity Measurement
JMIR Mhealth Uhealth 2021;9(3):e23391
URL: https://mhealth.jmir.org/2021/3/e23391
doi:10.2196/23391
PMID:33688843

Original Paper

Health Apps for Combating COVID-19: Descriptive Review and Taxonomy

Manal Almalki[1], PhD; Anna Giannicchi[2], MA

[1]Department of Health Informatics, Faculty of Public Health and Tropical Medicine, Jazan University, Jazan, Saudi Arabia

[2]Berkeley College, New York, NY, United States

Corresponding Author:
Manal Almalki, PhD
Department of Health Informatics
Faculty of Public Health and Tropical Medicine
Jazan University
Gizan
Jazan, 82726
Saudi Arabia
Phone: 966 597636563
Email: manal_almalki@hotmail.com

Abstract

Background: Mobile phone apps have been leveraged to combat the spread of COVID-19. However, little is known about these technologies' characteristics, technical features, and various applications in health care when responding to this public health crisis. The lack of understanding has led developers and governments to make poor choices about apps' designs, which resulted in creating less useful apps that are overall less appealing to consumers due to their technical flaws.

Objective: This review aims to identify, analyze, and categorize health apps related to COVID-19 that are currently available for consumers in app stores; in particular, it focuses on exploring their key technical features and classifying the purposes that these apps were designed to serve.

Methods: A review of health apps was conducted using the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines. The Apple Store and Google Play were searched between April 20 and September 11, 2020. An app was included if it was dedicated for this disease and was listed under the health and medical categories in these app stores. The descriptions of these apps were extracted from the apps' web pages and thematically analyzed via open coding to identify both their key technical features and overall purpose. The characteristics of the included apps were summarized and presented with descriptive statistics.

Results: Of the 298 health apps that were initially retrieved, 115 met the inclusion criteria. A total of 29 technical features were found in our sample of apps, which were then categorized into five key purposes of apps related to COVID-19. A total of 77 (67%) apps were developed by governments or national authorities and for the purpose of promoting users to track their personal health (9/29, 31%). Other purposes included raising awareness on how to combat COVID-19 (8/29, 27%), managing exposure to COVID-19 (6/29, 20%), monitoring health by health care professionals (5/29, 17%), and conducting research studies (1/29, 3.5%).

Conclusions: This study provides an overview and taxonomy of the health apps currently available in the market to combat COVID-19 based on their differences in basic technical features and purpose. As most of the apps were provided by governments or national authorities, it indicates the essential role these apps have as tools in public health crisis management. By involving most of the population in self-tracking their personal health and providing them with the technology to self-assess, the role of these apps is deemed to be a key driver for a participatory approach to curtail the spread of COVID-19. Further effort is required from researchers to evaluate these apps' effectiveness and from governmental organizations to increase public awareness of these digital solutions.

(JMIR Mhealth Uhealth 2021;9(3):e24322) doi:10.2196/24322

KEYWORDS

app; COVID-19; corona; self-care; personal tracking; review; mHealth; track; surveillance; awareness; exposure; consumer health informatics

Introduction

In response to the COVID-19 pandemic, a global health movement developed with countrywide campaigns providing health education. This information was widespread to educate the public on the newly discovered SARS-CoV-2 virus and how to best protect themselves. These campaigns were filled with a series of prevention protocols and control interventions to contain COVID-19, such as social distancing, keeping infected individuals isolated, self-isolating in homes or hotels after coming into contact with someone who has tested positive, and restricting travel [1].

During the COVID-19 pandemic, health mobile phone apps have been widely used for supporting these campaigns' missions, to assist in raising awareness on how the population may protect itself, and for encouraging adherence to those various precaution protocols [2,3]. For example, the United Kingdom National Health Service created an app with the purpose of encouraging users to keep a safe distance from others and alert them if they come close to someone who had been diagnosed with COVID-19. Within days, the app was downloaded over 10 million times, with six million downloads on the first day [4]. Globally, the World Health Organization (WHO) Alert app, a messaging service provided via social media outlets Facebook and WhatsApp that can be accessed in 15 different languages to answer questions about COVID-19, has the potential to reach two billion people [5].

In addition, health apps are not only limited to simply providing information about COVID-19 but also used to facilitate data-driven disease surveillance, screening, triage, diagnosis, and monitoring by governments or health officials, health care professionals, and health organizations [3,6]. These interventions enable timely preventative methods and treatment procedures at the population level [7]. For example, an app named COOPERA is used to monitor trends in COVID-19 in Japan, evaluate the current Japanese epidemiological situation, and provide useful insights to assist political decisions to tackle the epidemic. COOPERA collects personal information about the users and symptoms related to COVID-19. The reported number of confirmed infected cases are then calculated to detect and manage infectious cases.

At the individual level, as these health apps' popularity rises, the opportunity has increased for consumers and patients to self-manage both their risk of exposure and symptom progression [8,9]. In combatting COVID-19, health self-management includes keeping safe distance from others to decrease the communal spread of the disease, completing self-assessments that monitor symptom development or augmentation, routinely taking prescribed medication, and maintaining a healthy diet and physical activity [1,10]. In instances such as these, mobile apps can promote health self-care practice and activate a person's responsibility and accountability for preventing disease and maintaining health [9,11,12]. They can log and view the history of their health status, set useful reminders for treatment adherence, and provide vital information about their health status to the community to prevent future exposure.

Furthermore, the importance of health apps related to COVID-19 arises from their capabilities to allow consumers to feel safe and informed in making decisions regarding their health. For instance, an app developed by the University of California, San Francisco [13] for health care workers at the university's hospital assesses their potential COVID-19–related symptoms. This app helps the employees avoid long queues at screening points before each clinical shift, allows for physical distancing at hospital entrances, and prevents employees with suspected cases from coming to work. Moreover, it allowed the users to answer questions about symptoms they were experiencing, such as fever, cough, and difficulty breathing. The user can then self-assess their severity and make decisions about the potential need to seek further medical treatment [14].

To ensure that apps such as these can meet consumers' needs and preferences, it is imperative that app developers understand the various usages of these technologies and their technical features [8,15]. However, few studies have been conducted to describe COVID-19–related apps and analyze their technical features [15,16]. The lack of this understanding led developers and governments to make poor choices about health apps' designs, which led to creating less useful apps that are overall less appealing to consumers due to their technical flaws [17]. Therefore, this review aims to identify, analyze, and categorize health apps related to COVID-19 that are currently available for consumers in app stores; particularly, it focuses on exploring their key technical features and classifying the purposes that these apps were designed to serve.

Methods

Identification of Relevant Apps

From April 20, 2020, to September 11, 2020, the authors explored the Apple Store and Google Play. The following search strings were used to find apps dedicated for COVID-19 that were listed under the health and medical categories: COVID19, COVID, COVID-19, corona, coronavirus, corona triage, corona symptoms, SARS-CoV-2, and respiratory diseases. Additionally, the authors searched current news articles and Google search engine results to find apps related to COVID-19 that may not have otherwise been available in the authors' regional app store.

The app review was conducted with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines [18]. Although PRISMA-ScR is the standard guideline to review scientific literature, some researchers have suggested it is not completely applicable to app reviews [19,20]. Nevertheless, it has been used by several studies to review apps, as it is a common tool for performing systematic searches [19-21].

Screening and Eligibility Assessment

The inclusion and exclusion criteria are illustrated in Table 1 and described in light of a framework developed by Ramakrishnan et al [22]. The framework encompasses a series of key questions around COVID-19–related apps that are ordered by five levels of considerations, starting from wider and going to narrower considerations as the level's number gets higher. Although this framework was adapted from the M-Health Index and Navigation Database that has been previously used to evaluate mental health apps [22], in this study, it was used to help report our inclusion and exclusion criteria.

During the eligibility assessment round, the titles, descriptions, and keywords of identified apps were screened. Health apps that were available to the public at app stores were included. In this round, an app was excluded if it was removed from an app store by its developer during our specified search period, even if it was originally available at the beginning of this period. Their removal indicated that these apps were no longer available to consumers and were, hence, excluded from the sample.

An app was also excluded if it was only dedicated to respiratory or infectious diseases other than COVID-19, such as severe acute respiratory syndrome (SARS) or asthma without any reference to COVID-19. Examples of these apps were the Clean Your Lungs app [23], Breathcount [24], and Box Breathing [25]. Additionally, duplicated apps were identified and removed. In the event of a duplication, we included the Apple Store app, as it had all the required information about the app including its date of release.

Furthermore, no restrictions were made regarding the app's language, pricing, store location or country of origin, developer type, or accessibility measures for gaining access to its content, such as requirement of national identification codes, local country phone numbers, or research study codes. Additionally, no restrictions were imposed in terms of the type of app users. Examples of these health apps included Spectrum [26], which is intended for clinicians to help them in making clinical decisions based on COVID-19 guidelines and infection prevention protocols; Tabaud [27], which is intended for individuals who want to know if they have been in close proximity to an infected person; Self-quarantine [28], which is intended for patients or infected people who are in self-quarantine; Covid Radar [29], which is for researchers who want to predict health care needs in specific regions; and PreWorkScreen [30], which is for employers who want to manage their employee's COVID-19 self-screening.

Table 1. The inclusion and exclusion criteria of apps are presented based on a framework developed by Ramakrishnan et al [22].

Levels of consideration	Framework questions	Inclusion criteria	Exclusion criteria
App origin	Where does the app come from? Who is the developer, and what is the country of origin?	Health apps that were available at app stores during our search period	No restrictions were made on the country of origin, app's language, or developer type.
App accessibility	On what platforms is the app available? How much does it cost, and what accessibility measures are in place for a user?	Health apps that were available in the Apple Store and Google Play	No restrictions were made on the app's pricing or other accessibility measures for gaining access to the app's content such as requirement of national identification codes, local country phone numbers, or research study codes.
App features	What features does the app offer, and what kind of information is it providing around COVID-19?	Health apps that were related to combating COVID-19	Health apps were excluded if they were only dedicated to respiratory or infectious diseases other than COVID-19, such as severe acute respiratory syndrome or asthma without any reference to COVID-19.
Privacy and security	Are data use and security measures specified? What kind of data are collected or shared?	Health apps that were able to collect or share data	No restrictions were made on the privacy and security measures (eg, consent forms or privacy compliance standards) or kind of data collected or shared (eg, personal information or users' locations).
Clinical integration	For whom is it intended: patients, self-help, or essential workers?	Health apps that were available to the public	No restrictions were made on the app's intended users.

Selected Apps and Data Analysis

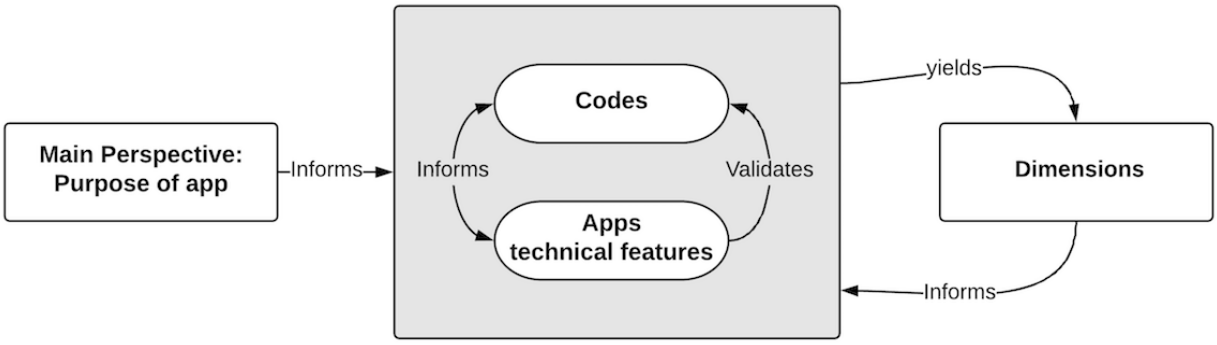
Each app's web page in both the Apple Store and Google Play was visited. Data on each app was extracted and collected as follows: the app's name, release date or version date when features related to COVID-19 were added, the country of origin, author or developer name, technical features, and source (link to the app's web page). This information is summarized and presented in Multimedia Appendix 1.

The apps' technical features were identified by performing open coding. A qualitative data analysis application (ie, Dedoose Version 8.3.35; SocioCultural Research Consultants, LLC) was used at this stage. The apps' descriptions were first stored in a Word (Microsoft Corporation) document, which was then imported into Dedoose. Excerpts about technical features were highlighted and given a title. The code title was drawn from the excerpts' content. For example, the excerpt "users can also engage in real-time chat with the chat feature" was placed under the title Chatbot Feature. Multimedia Appendix 2 presents the occurrence of all extracted excerpts in each app included in this review, whereas the excerpts' content can be found in Multimedia Appendix 1 within the column technical features.

After completion, the same process was repeated to identify the types of authors or developers creating these apps. Information about the authors or developers was summarized with some additional visits to their original websites to obtain more information.

After this, the generated codes about the apps' technical features were grouped into overarching categories or dimensions. Each dimension represented a different purpose that an app could serve. To identify and validate the purposes of these apps, codes and dimensions were compared iteratively to analyze the similarities in their descriptions within a category [31]. Author MA also compared the differences in codes in every other category as shown in Figure 1 (adapted from Judson et al [13]). If a code did not fit with the previously created dimensions, a new one was added. These generated dimensions were then summarized, which led to the development of our taxonomy of the apps' purposes (as shown in the Results section).

Figure 1. Our approach to defining purposes of apps related to COVID-19.

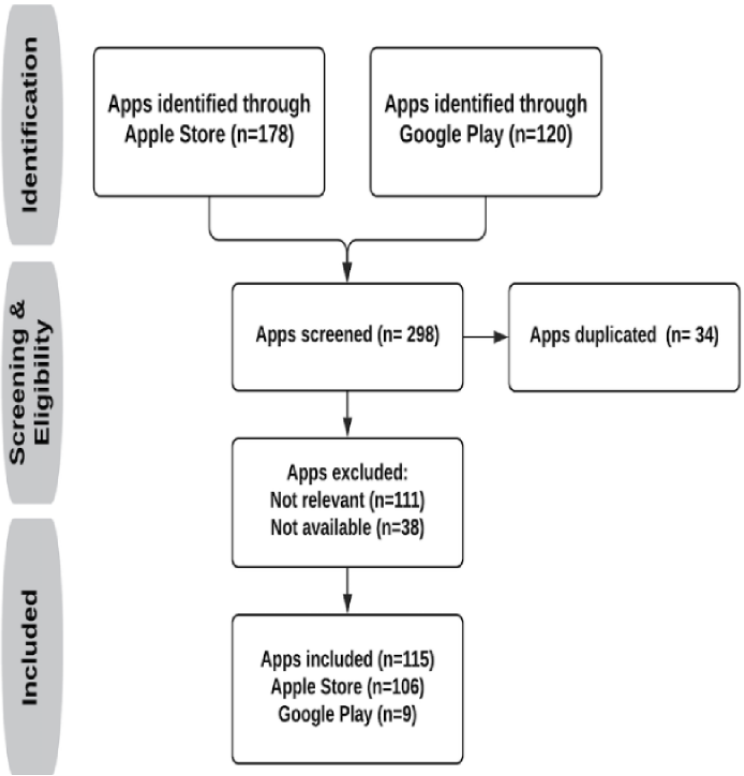


Results

App Selection

A total of 298 apps were identified through systematic searches in the Apple Store (n=178) and Google Play (n=120). Screening the apps' titles and descriptions resulted in removing the following apps: 34 apps that were duplicated, 111 apps that were related to other respiratory diseases (n=56 related to SARS and n=55 related to asthma), and 38 apps that were no longer available in the app stores. After removing duplicate and irrelevant apps, 106 apps from the Apple Store and 9 from Google Play were included and further analyzed. Figure 2 illustrates the flowchart of the search strategy and app selection process.

Figure 2. Flowchart of our search strategy and app selection process.



Characteristics of the Included Apps

The characteristics of included apps along with examples of each are presented in Table 2. Of the 115 health apps, 114 (99%) were free. Most of the apps (n=77, 67%) were developed by governments or national authorities such as the Ministry of Health in Saudi Arabia and the Government of Malaysia. Nearly one-fourth (n=25, 22%) of the apps were developed by companies such as BioneXt Lab, Paxera Health, and Webdoctor Limited. There were 5 apps from nonprofit organizations or agencies such as the WHO and Lay First Responders International. There were 4 apps built by universities or research centers such as Columbia University in the United States and the National Institute of Infectious Diseases in Romania. There were 4 apps developed by hospitals such as Merciful Brothers Hospital in Czechia.

Table 2. Summary of apps' characteristics.

Characteristics	Apps (N=115), n (%)	Selected app example
Operation system		
iOS	106 (92.17)	TraceCovid
Android	9 (7.82)	Tawakkalna KSA[a]
Pricing		
Free	114 (99.13)	Stop Covid19
Not free	1 (0.87)	PreMedicus
Developer/author		
Government or national authority	77 (66.95)	Covid-19 Armenia
Company	25 (21.73)	Covive
Nonprofit organization or agency	5 (4.35)	WHO[b] Info
Hospital	4 (3.48)	Trecovid19
University or research institute/center	4 (3.48)	Covid Watcher
The country of origin		
US	28 (24.35)	CovidWise
India	5 (4.35)	AarogyaSetu
Italy	4 (3.48)	Trecovid19
KSA	4 (3.48)	Tabaud
Mexico	4 (3.48)	Plan Jalisco Covid-19
Spain	4 (3.48)	GVA CoronVirus
Vietnam	4 (3.48)	Covid-19 Vietnam
Global	4 (3.48)	Covive
Canada	3 (2.61)	Canada Covid-19
Australia	3 (2.61)	MyAus Covid-19
France	3 (2.61)	Covidom Patient
Netherlands	3 (2.61)	COVID-19 Medisch Dossier
UAE	3 (2.61)	TraceCovid
Malaysia	2 (1.74)	MyTrace
Ireland	2 (1.74)	PatientMpower for COVID-19
South Korea	2 (1.74)	Self-Quarantine app
UK	2 (1.74)	NHS24 Covid-19
Other	35 (30.43)	N/A[c]

[a]KSA: Kingdom of Saudi Arabia.

[b]WHO: World Health Organization.

[c]N/A: not applicable.

A total of 51 countries were included in this review. Most of the health apps (n=28, 24%) were from the United States, and 5 apps came from India. The following countries each created 4 apps: Italy, the Kingdom of Saudi Arabia, Mexico, Spain, and Vietnam. The following countries each made 3 apps: Australia, Canada, France, the Netherlands, and the United Arab Emirates. The following countries each developed 2 apps: Malaysia, Ireland, South Korea, Switzerland, and the United Kingdom. There were 4 apps designed for global users, and the rest came from other countries (n=35; ie, Armenia, Austria, Bahrain, Bolivia, Brazil, Columbia, Czechia, Denmark, Egypt, Georgia, Hungary, Indonesia, Iceland, Jamaica, Jordan, Kuwait, Lativa, Mali, Morocco, New Zealand, North Macedonia, Pakistan, Poland, Portugal, Qatar, Republic of Cyprus, Republic of Fiji, Romania, Singapore, Sri Lanka, Switzerland, Thailand, Tunisia, Turkey, and Uruguay).

Technical Features of Apps Related to COVID-19

After conducting the open coding of 115 apps' descriptions, 258 extracted excerpts were grouped into 29 technical features. Table 3 presents the key technical features with examples of apps that supported these technical features, and Multimedia Appendix 2 illustrates the occurrence of the different technical features in each app. Each technical feature is described in detail in the following sections.

The most common technical feature was *basic health information and advice* or *frequently asked questions* (*FAQs*). Over one-third (n=42, 37%) of the apps were developed to provide basic health information about COVID-19, best health practices, medical advice, and FAQs regarding COVID-19. This health information was given in the form of guidance documents, videos, and animation clips that were curated by the respective country's government, the WHO, Centers for Disease Control and Prevention (CDC), the National Health Council, Johns Hopkins University, and other medical institutions.

The second most common technical feature in our review was *contact tracing*. Over one-fourth (n=32, 28%) of the apps supported contact tracing by documenting and retaining encounters with others such as friends, family, or coworkers. These apps allowed users to detect other devices with the same installed app and exchange an encrypted Secure Tracing Identifier (STI). The STI is stored locally on the user's device and consists of anonymized data, a time stamp, and (in some apps) the GPS location of the phone. When any one of the users becomes infected with COVID-19, authorities with authorized access to the data can request the infected users to upload the list of STIs to their national data centers for further analysis to enable officials to respond quickly and reach out to individuals who were in close contact and who may be requested to quarantine, thus potentially minimizing the spread of the virus.

The third most common technical feature was *alert contacts*. Over one-fourth (n=30, 26%) of the apps provided the opportunity to alert contacts. These apps enabled users diagnosed with COVID-19 to voluntarily share their tests results with people they had come into contact with over the previous 14 days. These alerts were received through notifications, text messages, and automatic calls. In contrast, some apps were configured to allow health officials to automatically and anonymously inform the user's contacts of any encounters; the users' consents were usually obtained during the initial download of the app.

The fourth most common technical feature was *gadget of self-assessment*. There were 20 (17%) apps that provided self-assessment tools to examine whether the user may have COVID-19. These tools included questionnaires that were designed according to CDC guidelines, WHO recommendations, and the country's health officials. Based on the results from these questionnaires, users were ultimately divided into three categories: users who had not tested positive for the virus and are asymptomatic; users who had not tested positive for the virus but had symptoms such as a fever, cough, or shortness of breath; and users who had tested positive for COVID-19. The collected information from these apps was then locally or remotely processed and then made available to the country's health care professionals for further analysis.

The fifth most common technical feature was *live statistics and rolling updates*. There were 19 (16.5%) apps that provided live statistics and rolling updates in two forms: Really Simple Syndication (RSS) feeds and push notifications. RSS feeds were used by 14 apps, as shown in Multimedia Appendix 2, to provide up-to-date information about the COVID-19 infection by number of active cases. The cases were then divided into asymptomatic, mild, moderate, severe, recovered, and fatal. These statistics were grouped and illustrated per day, per week, and per month from both a countrywide and worldwide perspective. Additionally, some apps presented the statistics based on hospital admissions within the country rather than active cases. In this same vein, 5 apps used push notifications to provide updates from the government and its advisories, and updated information and subsequent instructions informing of the COVID-19 spread aggregated by the user's state and country as well as around the world.

The technical feature *latest news* came next in popularity with 16 (14%) apps. A variety of content such as stories, videos, and podcasts were advertised to present the most current global and local news feeds and, in some apps, can be sorted by cities and countries. These apps allow their users to receive immediate COVID-19–related news and updates from trusted nonprofit groups, international organizations, and government agencies in one place.

The technical feature *symptoms tracker* was also supported by 16 (14%) apps. These trackers log symptoms and vitals such as fever and cough at specific frequencies by sending text messages and emails, providing automatic fill-ins for text fields, or sending push notifications to obtain the required data. The collected data from symptom trackers assists in categorizing individuals by severity and health risks into "low-risk" to "high-risk" groups. This method of grouping patients was then used to decide the health care needed for each case. For example, in low-risk symptoms groups, users were provided with knowledge and resources to deal with the disease at home. In higher-risk groups, patients were monitored in anticipation of developing symptoms.

The technical feature *information about health services and care lines* came with 11 apps offering this feature. The apps

allowed the users to search nearby hospitals, emergency services, pharmacies, and certified COVID-19 test labs. Some apps consisted of a step-by-step guide on finding testing services and centers that were available around the users' locations as well as providing the contact information of these services.

The technical feature *map* came with a total of 10 apps. These maps interactively presented both the occurrence of COVID-19 cases (eg, active, confirmed, and recovered cases) as well as the density of these cases in different areas (eg, neighborhoods, cities, and countries). Maps were also used to indicate nearby health care centers and route directions to reach these centers. Moreover, these apps helped health officials observe trends in the community and in turn take meaningful measures to handle the spread of the virus.

The technical feature *health or travel declaration* came with a total of 8 apps. These apps detailed health or travel declarations that were mandated by the country's government or health officials. Upon arrival, travelers were requested to report themselves and their families through these apps and record their daily health status for 14 days. Moreover, some of these apps were used to notify users of potential exposure risk in the area where the users lived.

The technical feature *location monitoring* came with 7 apps. These apps provided real-time dashboards that help in identifying the next potential COVID-19 hot spots and monitoring and advocating for resources needed in those spots. The collected data were analyzed based on the geolocation of cases and then used to provide support in coordination with local, departmental, and national authorities, which in turn assisted in planning optimal treatment delivery.

The technical feature *sharing data or story with others* was also supported by 7 apps. These apps have the capability to build health diaries describing the users' symptom development and allow consumers to share their COVID-19 stories with other users and on social networks like Facebook. The stories and diaries can be shared by posting text-based messages, recording voice-based messages, or uploading videos. The users were also able to share these diaries with medical personnel to receive a faster diagnosis.

The technical features *remote monitoring* and *virtual medical consultation* were both offered by 6 apps. Apps that were related to remote monitoring allowed health care professionals to monitor patients' health data such as heart rate and level of blood oxygen in real time. The generated data could then inform health care teams on how their patients lived on a daily basis and allow them to be immediately alerted if any patient needed critical medical attention. Apps from the virtual medical consultation category enabled virtual medical consultation, live video consultations, or bidirectional text-audio communications to provide personalized support between the users and their doctors.

The technical features *helplines* and *chatbots* were offered by 6 and 5 apps, respectively. Helplines connect users to consultants providing useful information related to COVID-19 and facilitate the introduction of patients to health workers over a toll-free number. Chatbots, on the other hand, are artificial intelligence (AI)–enabled agents who connect with patients through texting or a human-like voice. Chatbots offer personalized health advice via one-on-one conversations with users and help them find answers to their various questions in real time.

The technical feature *distance detection* was enabled by 3 apps. These apps were built to improve the user's ability to avoid close contact with other people around them. These apps can show how far away the user's device is from other devices within the same location.

The technical feature *recruitment of volunteers* was offered by 2 apps. These apps were designed to enable the recruitment of volunteers for conducting scientific studies and trials pertaining to COVID-19. These apps asked users to report their symptoms and other required information daily, with each submission generated into data for research purposes.

The technical feature *lists of products* was also supported by 2 apps. These apps present lists of products required for combating COVID-19 (eg, gowns, surgical masks, respirators, face shields, and hand sanitizers), and some allow for calculating the number of these products in stock to find the average consumption rate. This was useful for informing the users about these important products and helping to optimize the use of these resources.

Finally, the least common technical features in this review, each offered in only 1 app, were: checklists of surfaces that required disinfection, taking photos of surfaces, making medical appointments, medical check-up tracking (eg, diagnosis time or submission of diagnosis) during self-quarantine, medical report generators, medication tracking and reminders, mood tracking and mental status (eg, coping with stress), movement permits (eg, during curfew), results of a COVID-19 laboratory test, and using wearable devices for symptom tracking such as pulse oximeters for tracking oxygen saturation in the blood.

Table 3. Summary of the apps' technical features with examples.

Technical features (n=29)	Apps (N=115), n (%)	Selected example of app
Basic health info and advice or FAQs[a]	42 (36.52)	Covid-19 Czechia
Contact tracing	32 (27.83)	TraceCovid
Alert contacts	30 (26.08)	Tabaud
Gadget of self-assessment	20 (17.39)	Covive
Live statistics and rolling updates	19 (16.52)	NCOVI
Latest news	16 (13.91)	CDC[b]
Symptoms tracker	16 (13.91)	Corona Care
Info about health services and care lines	11 (9.57)	CoronApp-Colombia
Map	10 (8.69)	Corona Map
Health or travel declaration	8 (6.96)	Covid-19 Vietnam
Location monitoring	7 (6.08)	Private Kit: Safe Paths
Sharing data or story with others	7 (6.08)	Corona FACTS
Remote monitoring	6 (5.21)	Covidom Patient
Virtual medical consultation	6 (5.21)	Laziodr Covid
Helpline	6 (5.50)	Covid-19 UAE
Chatbot	5 (4.35)	HealthLynked COVID-19
Distance detection	3 (2.61)	VírusRadar
Recruitment of volunteers	2 (1.74)	Covid Radar
List of products for combating COVID-19	2 (1.74)	NIOSH[c] PPE[d] Tracker
Checklist of disinfected surfaces	1 (0.87)	Disinfection Checklist
Making medical appointments	1 (0.87)	GVA CoronVirus
Medical check-up tracking	1 (0.87)	Self-Quarantine app
Medical report generator	1 (0.87)	Premedicus
Medication tracking and reminders	1 (0.87)	Patientsphere for Covid19
Mood tracking and mental status	1 (0.87)	Covid Coach
Movement permits	1 (0.87)	Tawakkalna KSA[e]
Results of COVID-19 laboratory test	1 (0.87)	Tatamman
Taking photos of surfaces	1 (0.87)	Disinfection Checklist
Wearable devices for symptom tracking	1 (0.87)	PatientMpower for COVID-19

[a]FAQ: frequently asked question.

[b]CDC: Centers for Disease Control and Prevention.

[c]NIOSH: National Institute for Occupational Safety and Health.

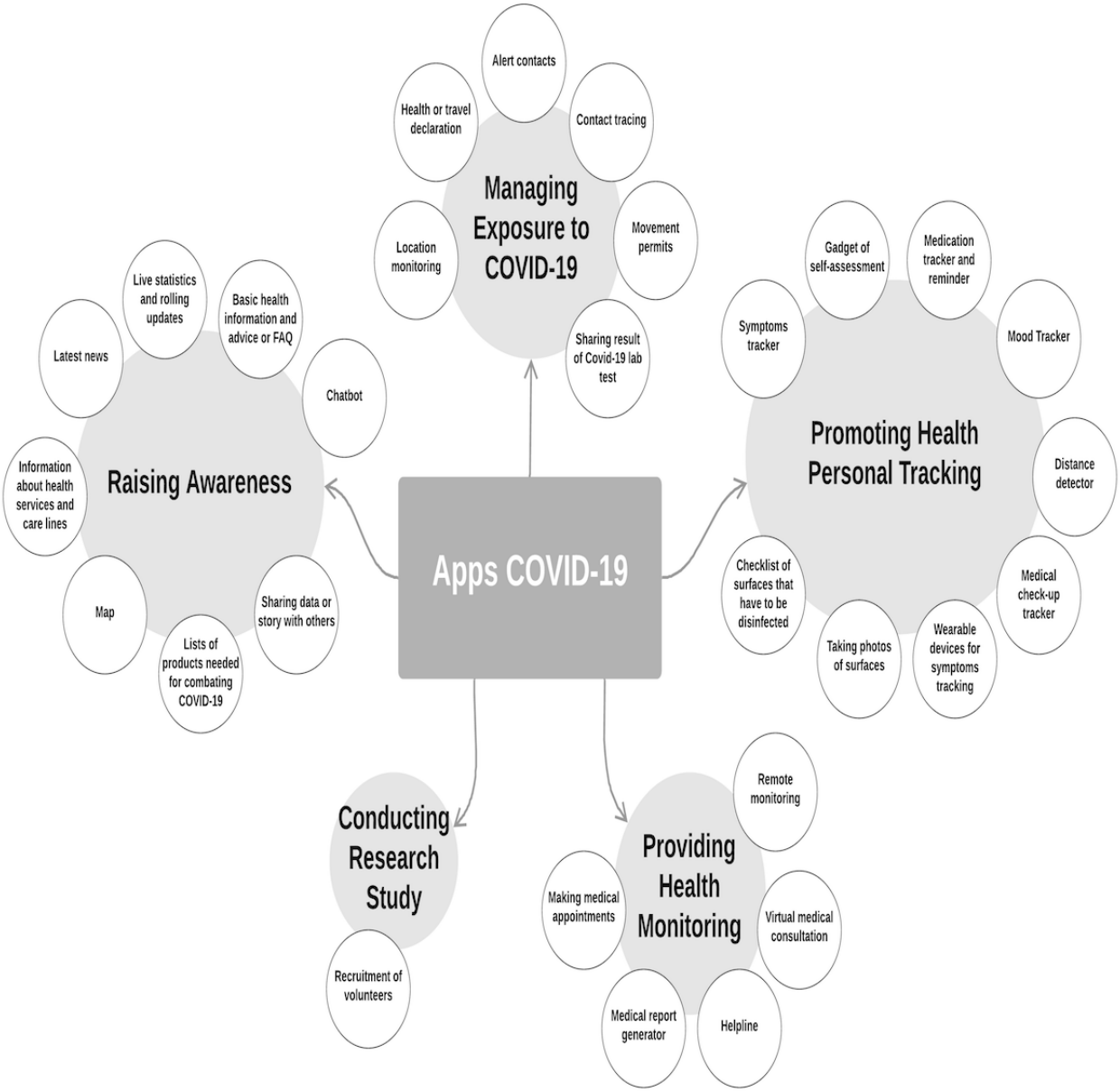
[d]PPE: personal protective equipment.

[e]KSA: Kingdom of Saudi Arabia.

Purposes of Apps Related to COVID-19

The identified technical features (n=29) were then analyzed and organized into five dimensions that represented purposes of these health apps. This led to the development of our taxonomy for health apps related to COVID-19, as shown in Figure 3. These purposes were tracking personal health, raising awareness, monitoring health by health care professionals, managing exposure to COVID-19, and conducting research studies. The majority of technical features were related to the first two purposes. Each purpose is described in the following section.

Figure 3. Taxonomy of COVID-19 apps' purposes. FAQ: frequently asked questions.



For tracking personal health, a total of 9 (31%) technical features out of 29 were found relevant, as the data generated from these apps helped their users look after their health. These technical features were gadgets of self-assessment used as an initial triage of possible infection, symptom trackers to check oneself for COVID-19 symptoms on a daily basis, medication trackers and reminders, mood trackers, distance detectors to maintain a safe distance from others, diagnosis recorders used during self-quarantine, integrated cameras for taking photos of surfaces that have been disinfected, checklists of surfaces that have to be disinfected for tracking hygiene practices, and wearable devices for tracking symptoms.

For raising awareness about COVID-19, 8 (28%) technical features out of 29 were categorized under this dimension, as these apps were concerned with providing various data and information to help users stay informed about this disease. These technical features included providing basic health information and advice or FAQs, presenting live statistics and rolling updates, showing the latest news, offering information about health services and care lines, providing interactive maps of active cases and nearby medical facilities to help users learn about this information, chatbots to answer the user's questions, incorporating lists of products needed for combating COVID-19 to learn about this important precaution, and allowing users to share data and stories with others such as family members and doctors to inform them about their health status and the most current health care information and practice.

For managing exposure to COVID-19, 6 (21%) technical features were classified under this dimension, as these apps help users avoid being exposed to this virus. These features were contact tracing, alerting users who were within close proximity of someone who had tested positive for the virus, reporting suspected cases and declaring travel after arrival, granting movement permits during curfews, monitoring the location of

consenting users to further understand trends of COVID-19 within various communities, and tools for sharing results of certified COVID-19 examinations with users.

For health monitoring, 5 (17%) technical features were placed under this dimension, as these apps help users to seek medical assistance from their health care professionals. These features were remote monitoring by health care professionals, virtual medical consultations with clinicians via video or audio calls, medical report generators, tools for making appointments to visit health centers, and helplines to obtain necessary medical help.

Finally, for conducting research studies, many features such as symptoms tracker, distance detector, and self-assessment gadgets enabled this dimension, but only 1 (3%) technical feature uniquely supported it. This feature was intended to enable researchers to recruit volunteers to take part in COVID-19–related studies and clinical research across countries.

Discussion

Principal Findings and Comparison to Prior Studies

This app review shows that the health apps related to COVID-19 vary in terms of their developer type, basic technical features, and purposes of use to combat COVID-19. Regarding the developer types, this review reveals the high number of apps developed by governments or national authorities to fight this infectious disease. In comparison, apps related to noncommunicable diseases like diabetes and hypertension are mostly developed by nongovernmental entities such as health care providers and hospitals [19,20]. This indicates the essential role that apps could play as powerful "weapons" in public health crisis management [32].

The most common technical features in this review focused on offering basic health information and advice on COVID-19 followed by contact tracing. However, the authors noticed that over the review period the number of health apps related to contact tracing was increasing. For example, health apps released at the early stages of the COVID-19 pandemic (ie, February, March, April, and May 2020) tended to be concerned with raising awareness more than other purposes. Apps released in June, July, August, and early September 2020 tended to offer COVID-19 exposure notifications more than any other purpose (see Multimedia Appendix 1). Moreover, it is also expected that contact tracing apps will continue to increase [33,34]. As COVID-19 is highly contagious, digital contact tracing is preferred by many health agencies and governments to expand their capacity for fighting the rapid spread of the virus [34,35]. Contact tracing via apps is also deemed more effective compared to traditional data collection [36]. In traditional contact tracing, tracing is performed through hired tracers and notifications are made by phone calls, which is costly and time consuming, and the chance of missing potential exposures is high [36].

This review also observed the increase in AI-based technical features for providing customized support to fight against the COVID-19 outbreak. For example, AI-enabled health education agents are offered through chatbots. Although 4 apps included in this review had chatbots within their technical features, the number of chatbot apps and their popularity are expected to increase [37]. Health chatbots can provide knowledge about the symptoms caused by the novel SARS-CoV-2 virus, teach their users how to detect if they have been affected by or infected with the virus, and present them with recommendations based on CDC guidelines and advice to reduce the risk of infection [10,38].

Moreover, some researchers are concerned about future lasting effects of the virus, which can be somewhat mitigated with currently available app technology [39]. For instance, Neubeck et al [40] conducted a review of the effectiveness of remote care to cardiovascular patients during the era of COVID-19. They found that there were two specific limitations that could not yet be answered by current mobile apps: rapidly changing symptoms and social isolation. This review found similar limitations in the current technology. Although chatbots are excellent in the initial stages of assessing symptoms or answering questions, they are not advanced enough to provide a sense of human contact that isolated or immunocompromised users may be interested in.

Regarding the purposes of apps, tracking personal health and raising awareness were dominant in this review and in other relevant reviews of COVID-19 apps [41,42]. As the COVID-19 outbreak has progressed, monitoring peoples' health and adherence to prevention procedures—including mobility, early detection of mild symptoms, and providing necessary psychological support—by health officials and national health care providers has become increasingly difficult to maintain and is anticipated to become even more challenging with an ever-increasing number of cases [43,44]. To combat these likely upcoming issues and at the request of health officials, many of the apps identified in this review had personal health-tracking capabilities. These apps support public health officials' emphasis on the importance of personal health tracking as a participatory approach to curtail further spread of COVID-19 in the community [45]. By involving most of the population in self-tracking their personal health and providing them with the technology to self-assess, the population can work in tandem with health care providers to combat the effects of the virus.

Additionally, in the 51 countries that are presented in this review, informing the community on how to mitigate their exposure to COVID-19 has been provided by digital health technologies [2]. The authorities in these countries and other stakeholders (eg, health care providers and policy makers) have used apps that provide services such as contact tracing and widespread alerts. Additionally, these apps assist in restricting travel and keeping contact with users, which shows the imperative need for these apps to keep the community safe. Likewise, as health care providers can use these apps to stay informed on the community's needs, they can in turn plan on how to best deliver treatment services. Each of these groups is responsible for their duty in combatting the virus; in using this technology, various stakeholders can feel confident in knowing what to do for reducing the risk of COVID-19 [2,6].

Implications of Findings

This study provides a holistic review of the technical features that are common in health apps related to COVID-19 and

identifies the most used ones for various purposes as shown in Figure 3. Knowledge of both the technical features and their applications in combating COVID-19 may help developers in making informed decisions when designing apps for various stakeholders.

For governments and health officials, our review shows not only that health apps can support several different methods in which to mitigate the effects of COVID-19 but also that its information can be accessed rapidly and inexpensively [6]. This review shows the ability of COVID-19–related apps to quickly and easily transmit data to both public health centers and treatment providers as well as epidemiologists, virologists, and clinicians. These health apps are equipped with different technical features such as contact tracing, where information can be uploaded and transmitted, privately held and maintained by health care authorities, yet still accessible to the public [46]. Individuals can use these apps to provide information about themselves or their social circle and can be alerted when approaching an area where the risk of exposure is high. These insights may encourage governments and health officials to request building apps in a way that can expand consumers' capabilities to feel educated, protected, and informed of their own health information and their community during pandemics.

For researchers in public health and medical informatics, there is enormous potential for future research and app-based development with regard to how mobile apps have so far been used to combat COVID-19, as shown in this review, and will ideally be used in the future [47]. One of the most frequent questions about these apps is related to their acceptability [48] and effectiveness [49]. These assessments, which are based on the apps' technical features (eg, the medication reminders, information on COVID-19 that the users can take advantage of), can inform not only successful individual adoption of these technologies but also effectiveness of overall services that apps can provide [47,50]. Our taxonomy of health apps purposes, illustrated in Figure 3, can be used as a guidance tool in categorizing apps and, assessing their functionalities, and evaluating their acceptability and effectiveness in each category.

Lastly, for average consumers, as health apps related to COVID-19 are low-cost and publicly available resources, this study provides a holistic reference of apps that are currently available in the market across 51 countries. Multimedia Appendix 1 is rich with information about the health apps' specifications and includes their webpages for convenient access. Multimedia Appendix 1 may help to educate consumers about various apps available and their different functionalities, which may improve their abilities to choose the best suitable app based on their needs and requirements [42].

Recommendations for Future Studies

As countries who are actively collecting data are better equipped to make decisions on how to best combat COVID-19, it is recommended to investigate and find approaches for encouraging as many people as possible to use these apps [17,50,51]. With an app, a user can at any time conduct a self-assessment to evaluate any symptoms they may have and take appropriate measures to keep themselves and their community safe. However, poor choices made by developers and governments about the apps' designs have led to technical flaws and security concerns that could make the apps less powerful and may hinder consumers' willingness to use them [17]. Therefore, these types of apps should be developed with health informatics experts to improve collaboration between government, health care organizations, and app developers, and to achieve the best quality of data collection and protection [32].

Self-monitoring–related apps have played a large role in health care provision, treating patients without being overwhelmed by in-person visits and enabling treatment providers and patients to use symptom progression tracking in real time [11]. However, health monitoring has to be distinguished as a concept from health tracking [52]. In health monitoring, the health care professionals, not users, take the initiative and provide guidance for their patients through the treatment course. In health tracking, the user takes initiative to complete actionable steps for health self-management [52]. Therefore, it is recommended that this difference be taken into consideration when the evaluation of app functionalities and applications is performed.

Additionally, although daily or even more frequent reminders of medication adherence, filling out self-assessments, and recommendations to self-quarantine may be effective for some users, others need more specialty care [53]. People who are at risk for rapidly changing symptoms must be informed that, when a new symptom suddenly develops, it is imperative that they immediately seek in-person treatment. Most of the apps included in this review did not explicitly have this type of technical feature. Although it has been important that these initially developed apps can answer the questions of a universal audience, there is a need for apps to be nationally accessible but exceptionally customizable.

Furthermore, one of the most pressing questions in clinical treatment and research is how to prevent the feelings of social isolation that quickly developed in a large portion of the population—particularly in older adult and immunocompromised populations [54]. There has been little research on the use of mobile apps to combat symptoms such as loneliness in a worldwide lockdown like the one enacted for COVID-19 [51]. Considering the individual and social effects of the virus that apps are not yet answering and in anticipation of future public health emergencies, medical experts and app developers looking to create innovative and useful products may want to consider amplifying a more personalized experience with more opportunity for human interaction. Future research can be conducted on both how feelings of depression, isolation, and loneliness may be reduced with the use of mobile health apps and if a personalized experience leads to beneficial, cost-effective in-person treatment [44].

Lastly, few health apps were used for conducting research studies. In this review, this was found to be the overall least common purpose of health apps related to COVID-19. Obtaining informative data on the novel SARS-CoV-2 virus from consumers is essential for public health specialists and medical researchers to successfully carry out their studies, which will inform future understanding of the infection and risk factors for adverse outcomes, characterize the virus transmission patterns, identify high-risk patients, and eventually assist clinicians in

fighting COVID-19 [43,55]. Therefore, further effort is required from governmental organizations to promote the conduct of participatory disease-based studies of the COVID-19 pandemic through the development of these digital solutions [45].

Limitations of This Study

This review has its own limitations. Our search for health apps was limited to the major app stores Apple Store and Google Play. However, these stores are the largest global platforms for app distribution, with 4.41 million apps as of May 2020, which accounted for about 80% of apps in the market [56]. Additionally, as searching all national app stores from a single country is difficult since international apps are not visible, other apps may have been missed [41,42]. Thus, other sources (eg, news and the Google search engine) were used to search for apps to overcome this problem. Furthermore, the quality of these apps was not examined and rated as suggested by Stoyanov et al [57], but rather, the apps' information that was retrieved from the app stores was thoroughly described and presented in Multimedia Appendixes 1 and 2. In addition, as 67% (77/115) of the included apps were developed by national health officials or governments, accessing some of these apps' content required national identification codes or local country phone numbers; this led the authors to making the decision to not perform an assessment of these apps' quality, as most information would be inaccessible [36,42].

Conclusions

This study provides an overview and taxonomy of the health apps currently available in the market to combat COVID-19 based on their differences in terms of basic technical features and purposes of use. The analysis of 115 health apps related to COVID-19 led to extracting 258 excerpts that were grouped into 29 technical features as shown in Table 3. These technical features were then categorized, which led to five overarching dimensions or purposes of apps for fighting COVID-19 as shown in Figure 3. Further effort is required from researchers to evaluate these apps' effectiveness and acceptability, and from governmental organizations to increase public awareness of these digital solutions. Our taxonomy of these apps' purposes can be used as a guidance tool in categorizing apps and then assessing their functionalities, as well as to evaluate their effectiveness in each category.

Acknowledgments

The authors are grateful to the Deanship of Scientific Research at Jazan University for funding this study (Project Cov19-12). We also would like to thank Ms Halah Abkar Zain, the lecturer at Jazan University, for her assistance during the processes of searching for apps and extracting excerpts related to the platform type, app name, pricing, name of the developer or author, country of origin, and technical features.

Authors' Contributions

MA conceived the study and acquired funding. MA designed the study, reviewed the apps, analyzed and interpreted the data, and wrote the first draft of the manuscript including the creation of all tables, figures, and multimedia appendixes. AG assisted in shaping the literature review with input from MA. Both authors contributed to the final version of the manuscript. Both authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary of the apps included in the review.
[PDF File (Adobe PDF File), 538 KB - mhealth_v9i3e24322_app1.pdf]

Multimedia Appendix 2
The occurrence of excerpts or technical features in each app.
[PDF File (Adobe PDF File), 531 KB - mhealth_v9i3e24322_app2.pdf]

References

1. COVID-19. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/index.html [accessed 2020-08-09]
2. Budd J, Miller BS, Manning EM, Lampos V, Zhuang M, Edelstein M, et al. Digital technologies in the public-health response to COVID-19. Nat Med 2020 Aug;26(8):1183-1192. [doi: 10.1038/s41591-020-1011-4] [Medline: 32770165]
3. Whitelaw S, Mamas MA, Topol E, Van Spall HGC. Applications of digital technology in COVID-19 pandemic planning and response. Lancet Digit Health 2020 Aug;2(8):e435-e440 [FREE Full text] [doi: 10.1016/S2589-7500(20)30142-4] [Medline: 32835201]
4. NHS COVID-19 app has been downloaded over 10 million times. GOV.UK. 2020. URL: https://www.gov.uk/government/news/nhs-covid-19-app-has-been-downloaded-over-10-million-times [accessed 2020-08-09]

5. WHO Health Alert brings COVID-19 facts to billions via WhatsApp. World Health Organization. 2020. URL: https://www.who.int/news-room/feature-stories/detail/who-health-alert-brings-covid-19-facts-to-billions-via-whatsapp [accessed 2020-08-09]
6. Alwashmi MF. The use of digital health in the detection and management of COVID-19. Int J Environ Res Public Health 2020 Apr 23;17(8):2906 [FREE Full text] [doi: 10.3390/ijerph17082906] [Medline: 32340107]
7. Yoneoka D, Kawashima T, Tanoue Y, Nomura S, Ejima K, Shi S, et al. Early SNS-based monitoring system for the COVID-19 outbreak in Japan: a population-level observational study. J Epidemiol 2020 Aug 05;30(8):362-370. [doi: 10.2188/jea.JE20200150] [Medline: 32475884]
8. Herendy C. How were apps developed during, and for, COVID-19?: an investigation into user needs assessment and testing. 2020 Presented at: 2020 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom); September 23-25, 2020; Mariehamn, Finland. [doi: 10.1109/coginfocom50765.2020.9237821]
9. Banskota S, Healy M, Goldberg E. 15 smartphone apps for older adults to use while in isolation during the COVID-19 pandemic. West J Emerg Med 2020 Apr 14;21(3):514-525 [FREE Full text] [doi: 10.5811/westjem.2020.4.47372] [Medline: 32302279]
10. Miner AS, Laranjo L, Kocaballi AB. Chatbots in the fight against the COVID-19 pandemic. NPJ Digit Med 2020;3:65. [doi: 10.1038/s41746-020-0280-0] [Medline: 32377576]
11. Leite H, Hodgkinson IR, Gruber T. New development: 'Healing at a distance'—telemedicine and COVID-19. Public Money Manage 2020 Apr 09;40(6):483-485. [doi: 10.1080/09540962.2020.1748855]
12. Kumar A, Gupta PK, Srivastava A. A review of modern technologies for tackling COVID-19 pandemic. Diabetes Metab Syndr 2020;14(4):569-573 [FREE Full text] [doi: 10.1016/j.dsx.2020.05.008] [Medline: 32413821]
13. Judson T, Odisho A, Young J, Bigazzi O, Steuer D, Gonzales R, et al. Implementation of a digital chatbot to screen health system employees during the COVID-19 pandemic. J Am Med Inform Assoc 2020 Jul 01;27(9):1450-1455 [FREE Full text] [doi: 10.1093/jamia/ocaa130] [Medline: 32531066]
14. Dennis A, Kim A, Rahimi M, Ayabakan S. User reactions to COVID-19 screening chatbots from reputable providers. J Am Med Inform Assoc 2020 Nov 01;27(11):1727-1731 [FREE Full text] [doi: 10.1093/jamia/ocaa167] [Medline: 32984890]
15. John Leon Singh H, Couch D, Yap K. Mobile health apps that help with COVID-19 management: scoping review. JMIR Nurs 2020;3(1):e20596 [FREE Full text] [doi: 10.2196/20596] [Medline: 32897271]
16. Jalabneh R, Zehra Syed H, Pillai S, Hoque Apu E, Hussein MR, Kabir R, et al. Use of mobile phone apps for contact tracing to control the COVID-19 pandemic: a literature review. SSRN J. Preprint posted online July 5, 2020. [doi: 10.2139/ssrn.3641961]
17. Servick K. Can phone apps slow the spread of the coronavirus? Science 2020 Jun 19;368(6497):1296-1297. [doi: 10.1126/science.368.6497.1296] [Medline: 32554573]
18. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
19. Keutzer L, Wicha SG, Simonsson US. Mobile health apps for improvement of tuberculosis treatment: descriptive review. JMIR Mhealth Uhealth 2020 Apr 21;8(4):e17246 [FREE Full text] [doi: 10.2196/17246] [Medline: 32314977]
20. Jamaladin H, van de Belt TH, Luijpers LC, de Graaff FR, Bredie SJ, Roeleveld N, et al. Mobile apps for blood pressure monitoring: systematic search in app stores and content analysis. JMIR Mhealth Uhealth 2018 Nov 14;6(11):e187 [FREE Full text] [doi: 10.2196/mhealth.9888] [Medline: 30429116]
21. Torous J, Keshavan M. COVID-19, mobile health and serious mental illness. Schizophr Res 2020 Apr;218:36-37 [FREE Full text] [doi: 10.1016/j.schres.2020.04.013] [Medline: 32327314]
22. Ramakrishnan AM, Ramakrishnan AN, Lagan S, Torous J. From Symptom Tracking to Contact Tracing: A Framework to Explore and Assess COVID-19 Apps. Future Internet 2020 Sep 08;12(9):153. [doi: 10.3390/fi12090153]
23. Clean your lungs. Google Play. URL: https://play.google.com/store/apps/details?id=com.dhadbadati.apps.clean_your_lung [accessed 2020-04-22]
24. Breathcount | asthma app. Google Play. URL: https://play.google.com/store/apps/details?id=lt.segfoltas.breathcount&hl=en [accessed 2020-04-22]
25. Box Breathing - Breath Air. Google Play. URL: https://play.google.com/store/apps/details?id=pl.orbitemobile.breathair&hl=en [accessed 2020-04-22]
26. Spectrum - Clinical Decision s. App Store. URL: https://apps.apple.com/sa/app/spectrum-clinical-decisions/id921339941 [accessed 2020-04-25]
27. Tabaud (COVID-19 KSA) . App Store. URL: https://apps.apple.com/sa/app/tabaud-covid-19-ksa/id1514704802 [accessed 2020-04-25]
28. 자가격리자 안전보호. Google Play. URL: https://play.google.com/store/apps/details?id=kr.go.safekorea.sqsm&hl=en [accessed 2020-04-25]
29. COVID Rada r. App Store. URL: https://apps.apple.com/sa/app/covid-radar/id1504947356 [accessed 2020-04-25]
30. Preworkscree n . App Store. URL: https://apps.apple.com/sa/app/preworkscreen/id1517912375 [accessed 2020-04-25]

31. Ohlin F, Olsson CM. Intelligent computing in personal informatics: key design considerations. In: Proceedings of the 20th International Conference on Intelligent User Interfaces. 2015 Presented at: IUI '15; March 2015; Atlanta, GA p. 263-274. [doi: 10.1145/2678025.2701378]
32. Ye J. The role of health technology and informatics in a global public health emergency: practices and implications from the COVID-19 pandemic. JMIR Med Inform 2020 Jul 14;8(7):e19866 [FREE Full text] [doi: 10.2196/19866] [Medline: 32568725]
33. Akinbi A, Forshaw M, Blinkhorn V. Contact tracing apps for COVID-19 pandemic: challenges and potential. OSF Preprints. Preprint posted online August 23, 2020. [doi: 10.31219/osf.io/6xbcs]
34. Raskar R, Nadeau G, Werner J, Barbar R, Mehra A, Harp G, et al. COVID-19 contact-tracing mobile apps: evaluation and assessment for decision makers. arXiv. Preprint posted online June 4, 2020 [FREE Full text]
35. Kleinman RA, Merkel C. Digital contact tracing for COVID-19. CMAJ 2020 Jun 15;192(24):E653-E656 [FREE Full text] [doi: 10.1503/cmaj.200922] [Medline: 32461324]
36. Leslie M. COVID-19 fight enlists digital technology: contact tracing apps. Engineering (Beijing) 2020 Oct;6(10):1064-1066 [FREE Full text] [doi: 10.1016/j.eng.2020.09.001] [Medline: 32953197]
37. Tudor Car L, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng Y, et al. Conversational agents in health care: scoping review and conceptual analysis. J Med Internet Res 2020 Aug 07;22(8):e17158 [FREE Full text] [doi: 10.2196/17158] [Medline: 32763886]
38. Li Y, Grandison T, Silveyra P, Douraghy A, Guan X, Kieselbach T, et al. Jennifer for COVID-19: an NLP-powered chatbot built for the people and by the people to combat misinformation. OpenReview. 2020. URL: https://openreview.net/pdf?id=HxIZzQZy_0F [accessed 2020-07-14]
39. Ribas RM, Campos PAD, Brito CSD, Gontijo-Filho PP. Coronavirus disease 2019 (COVID-19) and healthcare-associated infections: emerging and future challenges for public health in Brazil. Travel Med Infect Dis 2020;37:101675 [FREE Full text] [doi: 10.1016/j.tmaid.2020.101675] [Medline: 32305631]
40. Neubeck L, Hansen T, Jaarsma T, Klompstra L, Gallagher R. Delivering healthcare remotely to cardiovascular patients during COVID-19: a rapid review of the evidence. Eur J Cardiovasc Nurs 2020 Aug;19(6):486-494 [FREE Full text] [doi: 10.1177/1474515120924530] [Medline: 32380858]
41. 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]
42. 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]
43. Allam Z, Jones DS. On the coronavirus (COVID-19) outbreak and the smart city network: universal data sharing standards coupled with artificial intelligence (AI) to benefit urban health monitoring and management. Healthcare (Basel) 2020 Feb 27;8(1):46 [FREE Full text] [doi: 10.3390/healthcare8010046] [Medline: 32120822]
44. Mesa Vieira C, Franco OH, Gómez Restrepo C, Abel T. COVID-19: the forgotten priorities of the pandemic. Maturitas 2020 Jun;136:38-41 [FREE Full text] [doi: 10.1016/j.maturitas.2020.04.004] [Medline: 32386664]
45. Garg S, Bhatnagar N, Gangadharan N. A case for participatory disease surveillance of the COVID-19 pandemic in India. JMIR Public Health Surveill 2020 Apr 16;6(2):e18795 [FREE Full text] [doi: 10.2196/18795] [Medline: 32287038]
46. Fahey RA, Hino A. COVID-19, digital privacy, and the social limits on data-focused public health responses. Int J Inf Manage 2020 Dec;55:102181 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2020.102181] [Medline: 32836638]
47. Noronha N, D'Elia A, Coletta G, Wagner N, Archer N, Navarro T, et al. Mobile applications for COVID-19: a scoping review. Res Square. Preprint posted online April 29, 2020. [doi: 10.21203/rs.3.rs-23805/v1]
48. Fagherazzi G, Goetzinger C, Rashid MA, Aguayo GA, Huiart L. Digital health strategies to fight COVID-19 worldwide: challenges, recommendations, and a call for papers. J Med Internet Res 2020 Jun 16;22(6):e19284 [FREE Full text] [doi: 10.2196/19284] [Medline: 32501804]
49. Braithwaite I, Callender T, Bullock M, Aldridge RW. Automated and partly automated contact tracing: a systematic review to inform the control of COVID-19. Lancet Digit Health 2020 Nov;2(11):e607-e621 [FREE Full text] [doi: 10.1016/S2589-7500(20)30184-9] [Medline: 32839755]
50. Chidambaram S, Erridge S, Kinross J, Purkayastha S, PanSurg Collaborative. Observational study of UK mobile health apps for COVID-19. Lancet Digit Health 2020 Aug;2(8):e388-e390 [FREE Full text] [doi: 10.1016/S2589-7500(20)30144-8] [Medline: 32835196]
51. Saltzman LY, Hansel TC, Bordnick PS. Loneliness, isolation, and social support factors in post-COVID-19 mental health. Psychol Trauma 2020 Aug;12(S1):S55-S57. [doi: 10.1037/tra0000703] [Medline: 32551762]
52. Almalki M, Gray K, Martin-Sanchez FJ. Refining the concepts of self-quantification needed for health self-management. A thematic literature review. Methods Inf Med 2017 Jan 09;56(1):46-54. [doi: 10.3414/ME15-02-0007] [Medline: 27523820]
53. 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]

54. Kato TA, Sartorius N, Shinfuku N. Forced social isolation due to COVID-19 and consequent mental health problems: lessons from hikikomori. Psychiatry Clin Neurosci 2020 Sep;74(9):506-507 [FREE Full text] [doi: 10.1111/pcn.13112] [Medline: 32654336]
55. Lin L, Hou Z. Combat COVID-19 with artificial intelligence and big data. J Travel Med 2020 Aug 20;27(5) [FREE Full text] [doi: 10.1093/jtm/taaa080] [Medline: 32437541]
56. Number of apps available in leading app stores as of 1st quarter 2020. Statista. URL: https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/ [accessed 2020-08-09]
57. 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]

Abbreviations

AI: artificial intelligence
CDC: Centers for Disease Control and Prevention
FAQ: frequently asked questions
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
RSS: Really Simple Syndication
SARS: severe acute respiratory syndrome
STI: Secure Tracing Identifier
WHO: World Health Organization

Edited by L Buis; submitted 14.09.20; peer-reviewed by T Alessa, M Lozano-Lozano; comments to author 06.11.20; revised version received 09.12.20; accepted 22.02.21; published 02.03.21.

Please cite as:
Almalki M, Giannicchi A
Health Apps for Combating COVID-19: Descriptive Review and Taxonomy
JMIR Mhealth Uhealth 2021;9(3):e24322
URL: https://mhealth.jmir.org/2021/3/e24322
doi:10.2196/24322
PMID:33626017

Original Paper

Understanding On-Campus Interactions With a Semiautomated, Barcode-Based Platform to Augment COVID-19 Contact Tracing: App Development and Usage

Thomas Foster Scherr[1], PhD; Austin N Hardcastle[1], BE; Carson Paige Moore[1*], BS; Jenna Maria DeSousa[1*], BS; David Wilson Wright[1], PhD

Department of Chemistry, Vanderbilt University, Nashville, TN, United States

[*]these authors contributed equally

Corresponding Author:
Thomas Foster Scherr, PhD
Department of Chemistry
Vanderbilt University
7330 Stevenson Center
Station B 351822
Nashville, TN, 37235
United States
Phone: 1 6153222861
Email: thomas.f.scherr@vanderbilt.edu

Abstract

Background: The COVID-19 pandemic has forced drastic changes to daily life, from the implementation of stay-at-home orders to mandating facial coverings and limiting in-person gatherings. While the relaxation of these control measures has varied geographically, it is widely agreed that contact tracing efforts will play a major role in the successful reopening of businesses and schools. As the volume of positive cases has increased in the United States, it has become clear that there is room for digital health interventions to assist in contact tracing.

Objective: The goal of this study was to evaluate the use of a mobile-friendly app designed to supplement manual COVID-19 contact tracing efforts on a university campus. Here, we present the results of a development and validation study centered around the use of the MyCOVIDKey app on the Vanderbilt University campus during the summer of 2020.

Methods: We performed a 6-week pilot study in the Stevenson Center Science and Engineering Complex on Vanderbilt University's campus in Nashville, TN. Graduate students, postdoctoral fellows, faculty, and staff >18 years who worked in Stevenson Center and had access to a mobile phone were eligible to register for a MyCOVIDKey account. All users were encouraged to complete regular self-assessments of COVID-19 risk and to key in to sites by scanning a location-specific barcode.

Results: Between June 17, 2020, and July 29, 2020, 45 unique participants created MyCOVIDKey accounts. These users performed 227 self-assessments and 1410 key-ins. Self-assessments were performed by 89% (n=40) of users, 71% (n=32) of users keyed in, and 48 unique locations (of 71 possible locations) were visited. Overall, 89% (202/227) of assessments were determined to be low risk (ie, asymptomatic with no known exposures), and these assessments yielded a CLEAR status. The remaining self-assessments received a status of NOT CLEAR, indicating either risk of exposure or symptoms suggestive of COVID-19 (7.5% [n=17] and 3.5% [n=8] of self-assessments indicated moderate and high risk, respectively). These 25 instances came from 8 unique users, and in 19 of these instances, the at-risk user keyed in to a location on campus.

Conclusions: Digital contact tracing tools may be useful in assisting organizations to identify persons at risk of COVID-19 through contact tracing, or in locating places that may need to be cleaned or disinfected after being visited by an index case. Incentives to continue the use of such tools can improve uptake, and their continued usage increases utility to both organizational and public health efforts. Parameters of digital tools, including MyCOVIDKey, should ideally be optimized to supplement existing contact tracing efforts. These tools represent a critical addition to manual contact tracing efforts during reopening and sustained regular activity.

(JMIR Mhealth Uhealth 2021;9(3):e24275) doi:10.2196/24275

KEYWORDS

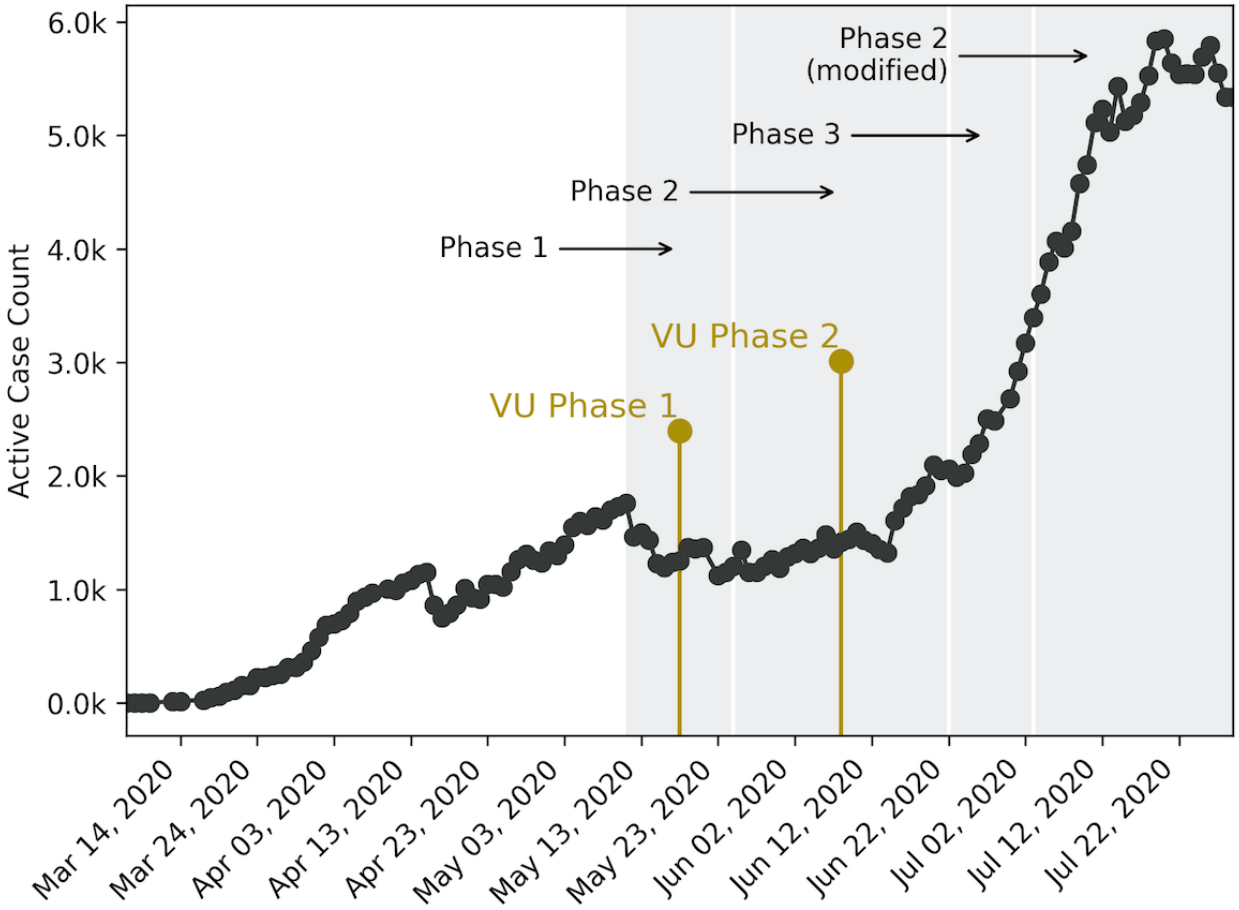
contact tracing; COVID-19; disease surveillance; digital health

Introduction

SARS-CoV-2, the virus that causes COVID-19, first emerged in late 2019. Months into the pandemic, the spread of COVID-19 continues to affect the world at large [1,2]. In response to COVID-19, entire countries enacted sweeping measures both nationally and in local hot spots. While these actions varied from country to country; in the United States, the declaration of a public health emergency led many state and local governments to implement "stay-at-home" directives, among other guidelines [3-6]. The ramifications were felt on state, city, and community levels; the consequences of these decisions included the closing of many nonessential businesses and a shift to remote work for many employees. Similarly, universities across the country closed research laboratories, removed undergraduate students from campus, and transitioned to virtual classrooms.

In Nashville, TN, the local government laid out a phased reopening of the city after the end of a stay-at-home order, which extended beyond the restrictions at the state level [7]. Phase 1, which began on May 11, allowed retail stores, restaurants, and bars serving food to open at 50% capacity, while high-touch and high-contact businesses such as nail salons, gyms, and entertainment venues remained closed. In phase 1, the Nashville Metro government encouraged social distancing and recommended, but did not require, face masks. Nashville's phase 2 of reopening began on May 25, increasing restaurant and retail capacity to 75% and opening high-touch businesses and entertainment venues at 50% and limited capacity, respectively. On June 22, Nashville entered phase 3 of the Metro reopening plan, although the city rolled back into a modified phase 2 stage on July 3 after a spike in cases (Figure 1) [8].

Figure 1. Active COVID-19 cases in Davidson County, TN, from mid-March through July 2020. Gray shaded boxes indicate the phases of the Nashville Metro Government reopening plan, while the gold lines indicate the start date of each phase of Vanderbilt University's (VU) reopening plan.



At Vanderbilt University, similar phased reopening steps were taken [9]. Each phase on campus mandated social distancing and masks, utilized on-campus pedestrian traffic plans, and encouraged remote work from staff or students when possible. The university entered phase 1 of their reopening on May 18, allowing research activities to resume at 33% capacity. On June 8, the university entered phase 2, allowing research capacity to increase from 33% to 50%, provided that 6 feet of social distance could be maintained between workers or students.

As states across the country begin to relax their precautionary measures and resume educational activities in the fall, it is generally understood that there is a need for increased vigilance and precautionary steps [10-12]. Many organizations are utilizing symptom-tracking software to monitor their community members during the reopening process, including in workplaces and on college campuses. Many freely available risk assessments have been widely distributed by public health entities, for-profit technology companies, and for-profit health care systems. While these are useful as informational tools and for understanding health disparities, there are concerns over the accuracy and utility of self-report symptom trackers in reopening efforts given the high degree of asymptomatic transmission associated with the current pandemic [13-16]. This highlights the need for other tools to focus on how to limit the spread from unknown transmission events.

Contact tracing has been a necessary method of identifying potential exposure events and understanding the epidemiology of the novel virus [17-24]. However, months into the pandemic, contact tracing remains largely a manual and labor-intensive process in which health care workers interview confirmed-positive COVID-19 incident cases and gather information on exposed people and locations. As case volumes grow and manual efforts struggle to handle the increase, it is clear that digital technology could assist with this process [25-28]. For instance, Apple and Google have partnered on a passive system that utilizes Bluetooth signals on mobile devices to identify when users are within a given distance for a certain time (a “contact event”) [27]. While Apple and Google have implemented best-in-class enhanced security features (eg, decentralized storage, rotating keys), security vulnerabilities have been identified in other strategies that rely exclusively on Bluetooth signals without similar protections in place [29-32]. Others have developed similar systems that utilize continuous GPS monitoring [25]. These approaches have raised substantial data ownership and privacy concerns, and early reports suggest that Bluetooth and GPS may struggle to accurately identify true contact through walls or on different floors of the interior floorplans common to office buildings and college campuses [33-38].

In response to these concerns, we developed MyCOVIDKey as an alternative digital contact tracing tool based on a combination of recurring risk assessments and a location check-in strategy. Since it relies on discrete event monitoring rather than continuous location monitoring or potentially vulnerable Bluetooth broadcasts, this approach is an alternative to current strategies and can provide an automated solution to supplement manual contact tracing efforts. The key-in feature of MyCOVIDKey, where users scan a location-specific barcode, can, importantly, augment existing contact tracing efforts in the face of asymptomatic transmission or inaccurate and unreliable symptom assessments. In this paper, we describe an app viability study in which we sought to understand the usefulness of this platform, its potential efficacy, and the sensitivity of its parameters.

Methods

Study Design

The Stevenson Center Science and Engineering Complex (Stevenson Center) of Vanderbilt University’s campus in Nashville, TN, was chosen as the study setting. Stevenson Center consists of 8 buildings in close proximity to one another. The buildings contain classrooms, research and teaching laboratories, graduate student and faculty offices, an engineering library (closed for the duration of the pilot study), and departmental administration offices. The buildings all have multiple floors, dedicated entrances and exits, stairwells, and elevators; several of the buildings are interconnected. For these reasons, Stevenson Center makes an ideal proxy for campuses at large, as well as moderately sized office complexes.

Laminated flyers (Figure 2C) were fixed to walls near building, stairwell, and elevator entrances, as well as in most common rooms and laboratories where users were expected to have returned to once on campus (Figure 3). Each flyer contained a barcode with a data payload of a unique hash code specific to that particular location. We elected to use PDF417 barcodes, commonly used on identification cards, instead of more common barcode types (ie, QR [quick response] code, data matrix). We believed that selecting a less common barcode that is not typically used to encode web addresses would have a positive impact on security by avoiding barcode hijacking (where a barcode is covered by another barcode that redirects a user to a malicious website) and requiring users to use our app instead of their mobile devices’ native camera app (most of which do not natively decode PDF417 barcodes). In total, there were 71 coded locations throughout the different buildings.

Figure 2. (A) The landing page of MyCOVIDKey, shown after a successful login. (B) A pop-up modal window that enables users to key in by scanning a location's bar code flyer. (C) A representative key-in flyer, with a barcode that has a unique embedded hash code specific to a location on campus.

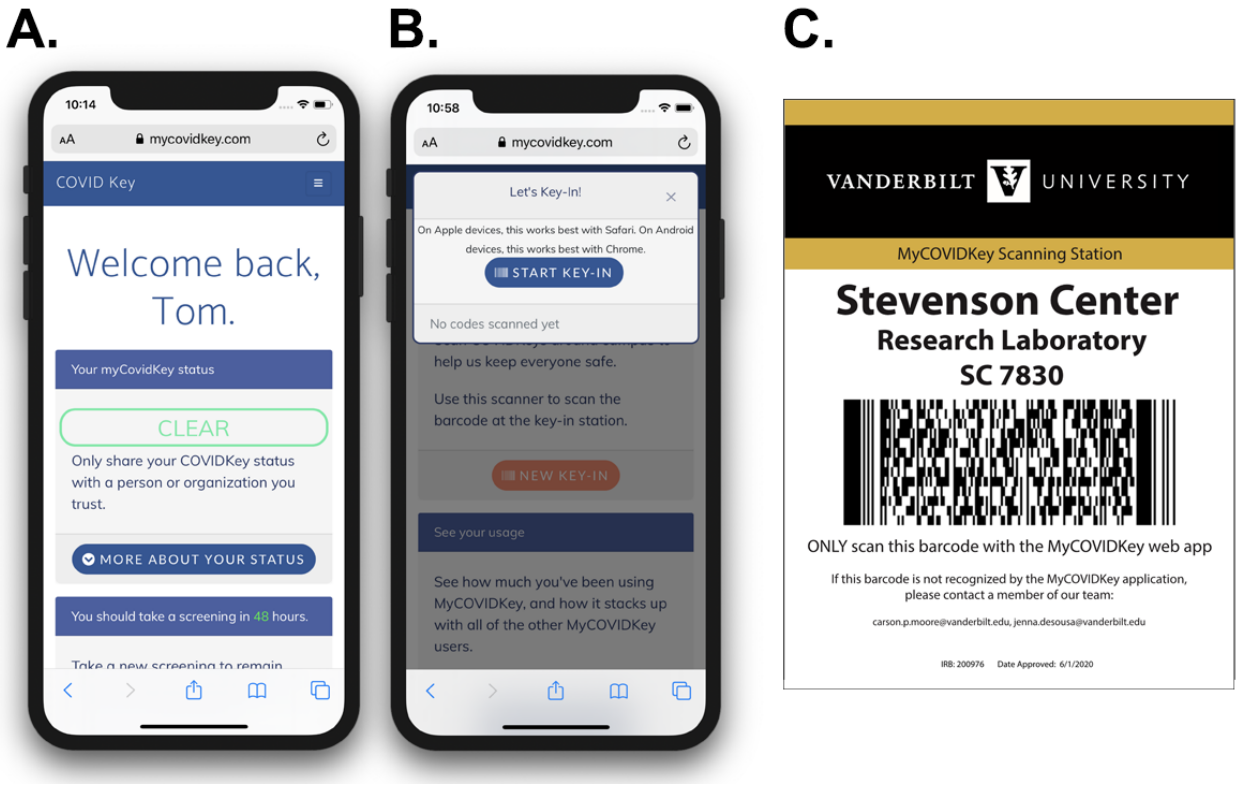
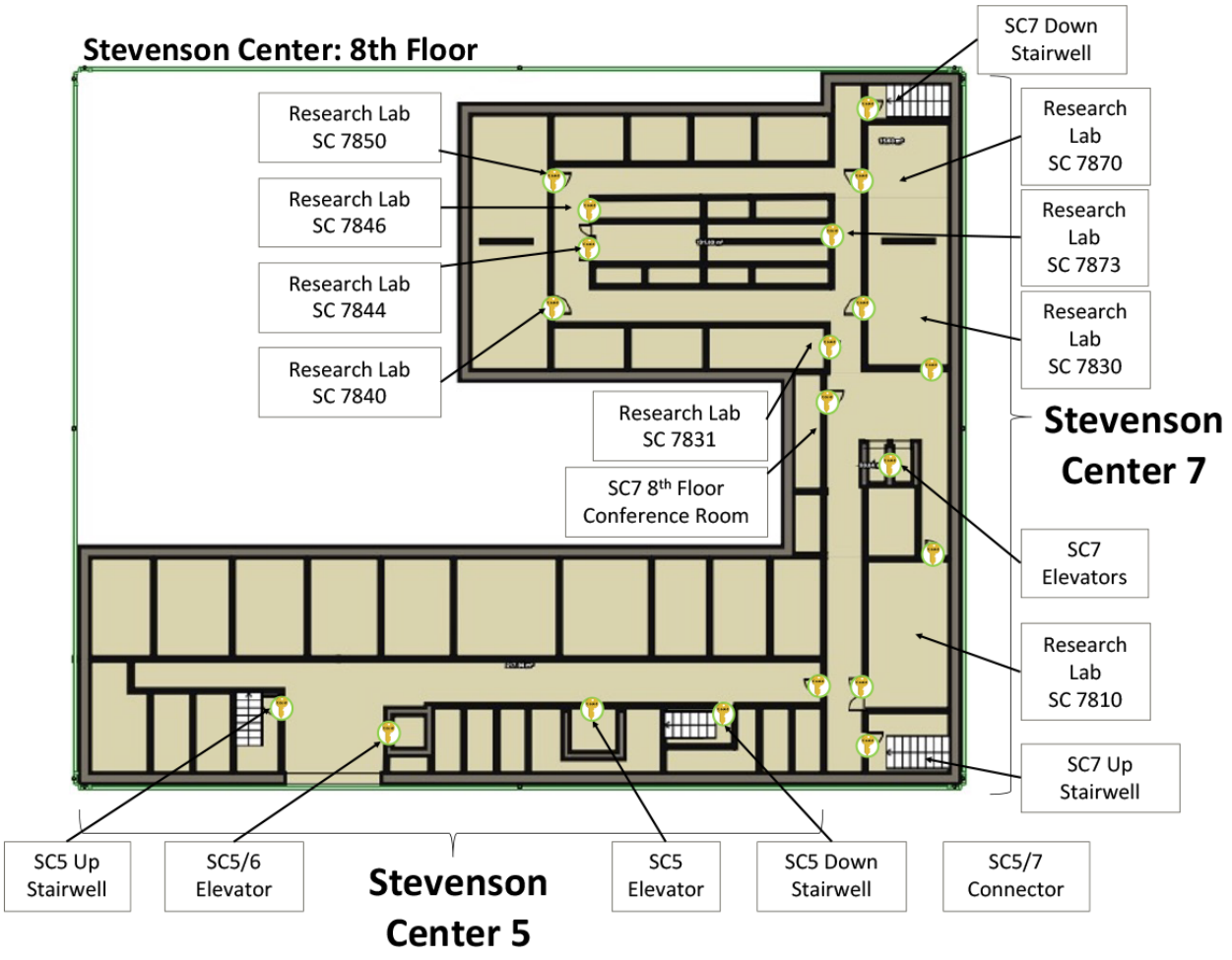


Figure 3. A coverage map of key-in flyers on the 8th floor of Stevenson Center (SC) 5 and Stevenson Center 7.



The study was set for 6 weeks and began on June 17, 2020. Participants were recruited via flyers posted throughout Stevenson Center as well as department-wide email lists. Users were provided brief instructions via a guided walk-through of the app the first 4 times that they arrived at the home screen. A weekly raffle based on usage was put in place as an incentive; however, all users were free to use the app at will. Upon completion of the pilot study, a survey was sent to all participants. This survey included questions about user demographics, as well as satisfaction questions focused on the MyCOVIDKey user experience. This work focuses on the technical implementation and results from the pilot; a thorough analysis of the postpilot survey, as well as a usability analysis and recommendations for improvement, are described elsewhere [39].

This study was reviewed and approved by the Vanderbilt University Institutional Review Board (#200976) on June 1, 2020.

App Design and Use

The MyCOVIDKey app was hosted by Amazon Web Services [40]. The platform consists of an Apache HTTP web server, a MySQL database, a custom-built PHP application programming interface, and a responsive, mobile-friendly (JavaScript, CSS, HTML) frontend. All data transmission between the server and client devices used secure protocols (HTTPS/SSL). A custom-built paradata capture library was included to perform usage analytics.

The app has a user hierarchy that includes specific privileges for four different classes of users: users, app administrators, contact tracers, and developers (Multimedia Appendix 1). All created accounts are users by default, with additional privileges accessible only if they have been granted by someone with the higher privilege level. With this structure, app administrator and contact tracer are distinct roles: the former sets parameters for use within the app but does not access user data; the latter performs the actual contact tracing with access to identifying information. This user hierarchy builds a foundation for enhanced privacy features where identifying user data can be siloed from deidentified but linked key-in and symptom information. Such an approach, which will be integrated prior to a wider rollout, would follow the lead set by the Apple/Google platform by saving different pieces of data on isolated servers that are managed by distinct user classes. Only in the event of a positive test will the user hierarchy coordinate to access the data necessary for contact tracing.

During account creation, participants provided an email address, password, phone number, name, birth date, and home zip code. Demographic data (age, sex, race) were not collected from users upon creation of a MyCOVIDKey account. After a successful login, users were directed to the landing page (Figures 2A and 4, and Multimedia Appendix 2). On this screen, separate tiles (Figure 5) could be expanded to display information on the user's current MyCOVIDKey status (including recommendations based on their most recent self-assessment), start a new self-assessment, present a modal window to perform barcode scanning at MyCOVIDKey locations, compare an individual's usage statistics to the entire cohort, and display their progress for the weekly raffle.

Figure 4. The home screen of MyCOVIDKey displays information about the user's current MyCOVIDKey status, allows users to perform self-assessments, key in to new locations, and view simple usage statistics. Certain features are disabled, and the text is adjusted to reflect a user's current status: (A) CLEAR, (B) NOT CLEAR.

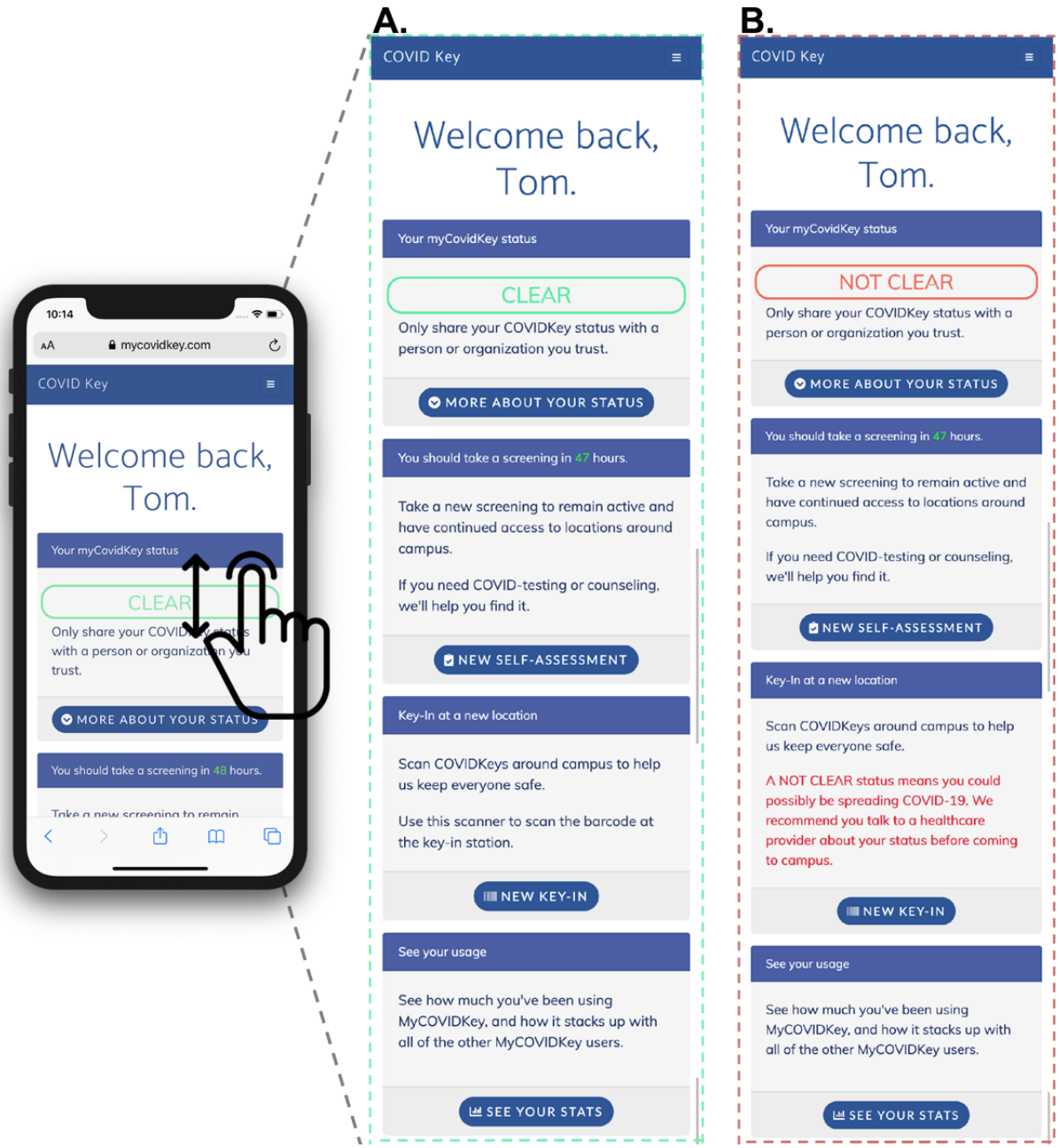
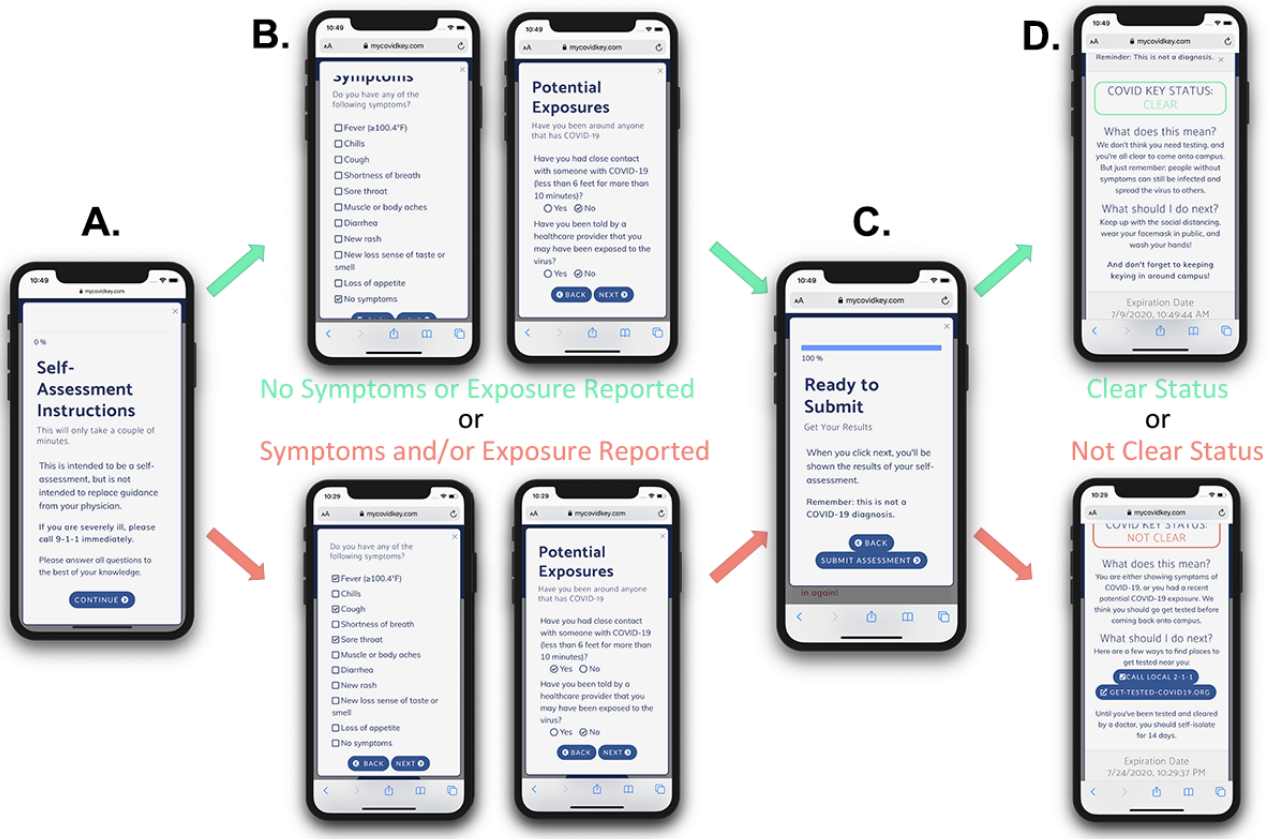


Figure 5. Recommendations were customized based on the user's current status: (A) NO STATUS, (B) NOT CLEAR, (C) CLEAR, and (D) EXPIRED.

The self-assessment was designed to be brief, since it was intended to be used repetitively, yet included COVID-19 symptoms outlined by the Centers for Disease Prevention and Control (CDC), as well as two questions designed to determine exposure risk. Symptom- and exposure-free users were given a status of CLEAR while a selection of any symptom or exposure would designate a status as NOT CLEAR (Figure 6). Although the user-facing result of the self-assessment was binary, internally self-assessments were coded using a point-based system to classify results as "low," "moderate," or "high." Our scoring system counted canonical COVID-19 symptoms (fever, chills, cough, and shortness of breath) and known exposure risks as 3 points; the presence of a rash or loss of smell and/or taste counted as 2 points; and a sore throat, body aches, and diarrhea were scored as 1 point. After summing the individual point values, the risk score was classified as follows: 0 points was defined as low risk, greater than 0 but less than 3 was defined as moderate risk, and greater than or equal to 3 was defined as high risk. While there are many ongoing efforts to distill qualitative COVID-19 symptoms to a numerical risk score, there currently is no standard approach for doing so. As such, the scoring system that we adopted proved useful to numerically differentiate users with canonical symptoms of COVID-19 from those with less specific symptoms.

Figure 6. The modal window to perform a self-assessment shows (A) brief instructions, (B) common symptoms and exposure risks of COVID-19, (C) a confirmation/submission screen, and (D) customized results based on the outcome of the self-assessment. Potential pathways to CLEAR and NOT CLEAR statuses are shown on top (green) and bottom (red), respectively.



Users with a CLEAR status were provided social support and encouragement to stay vigilant; those that received a NOT CLEAR status were instructed that the self-assessment was not a diagnosis and that they should seek diagnostic testing prior to returning to campus. The latter group was provided with a link to locate testing resources based on the zip code that they provided when their account was created [41]. When a self-assessment was completed, the user ID, symptoms, potential exposures, and the time stamp of the self-assessment were recorded. For this study, assessments were given an expiration of 48 hours, after which the key-in feature of the app was disabled until the user took a new self-assessment. This duration was chosen to increase the likelihood of continued usage by minimizing the burden on users during the pilot. However, the frequency of recurring self-assessments could readily be customized by organizational administrators to meet their needs. Upon completing a new self-assessment, the key-in feature was reactivated.

When a user entered a location with a key-in flyer, they could click the "New Key-In" button on the home screen to launch the key-in modal window. From there, the user was prompted to press the "Start Key-In" button, which initiated the barcode scanner (using the Scandit Software Development Kit, v5.0-5.1). When a user scanned a barcode, the app collected that event in the database, recording the user ID of the scanner, the time stamp of the scan, and the location ID that was scanned.

A weekly raffle was implemented on June 23 to incentivize participation. Users were allowed to accumulate entries in the drawing based on the number of self-assessments they performed and their number of key-ins each week. The number of entries was weighted for each event: each self-assessment was worth 10 entries in the raffle, and each key-in was worth 1 entry in the raffle. To avoid attempts to manipulate raffle outcomes by increased usage, the maximum number of entries a user could receive for each type of event was limited to 30.

Administrator features were included that allowed the study team to visualize usage metrics on a dashboard, perform manual contact tracing queries, and see results from the automated contact tracing algorithm. This algorithm is visually depicted in Multimedia Appendix 3. Briefly, when a participant completes a self-assessment that indicates either symptoms of or potential exposure to COVID-19, that creates a "person of interest" (POI) case. A case window is created that extends 48 hours prior to the causative self-assessment time stamp (the reverse case window) and continues for 14 days after the self-assessment (the forward case window). Any locations that the user keys in to during this period become "locations of interest." A second window of ±30 minutes is then created, centered around the time stamp of the POI's key-in at a particular location (the "contact overlap window"). Any other users who key in to the same location during the overlap window are deemed "contacts of interest." It is important to emphasize that these criteria are not the same as the CDC's guidelines for "close contact"; instead, our approach aligns with the goal of streamlining manual contact tracing efforts, rather than replacing them. As such, the lengths of the forward case window, the

reverse case window, and the contact overlap window can be customized based on organizational rules, manual contact tracing infrastructure and bandwidth, and location type.

Data Analysis

The data that were collected consisted of user information, the results of recurring self-assessments, data from key-ins, as well as app (usage) paradata. At the conclusion of the 6-week pilot, the data were exported from the database for analysis using Python statistical and visualization packages (Python Software Foundation). The data were then coded, identifiers removed, and then loaded into a REDCap project for long-term storage.

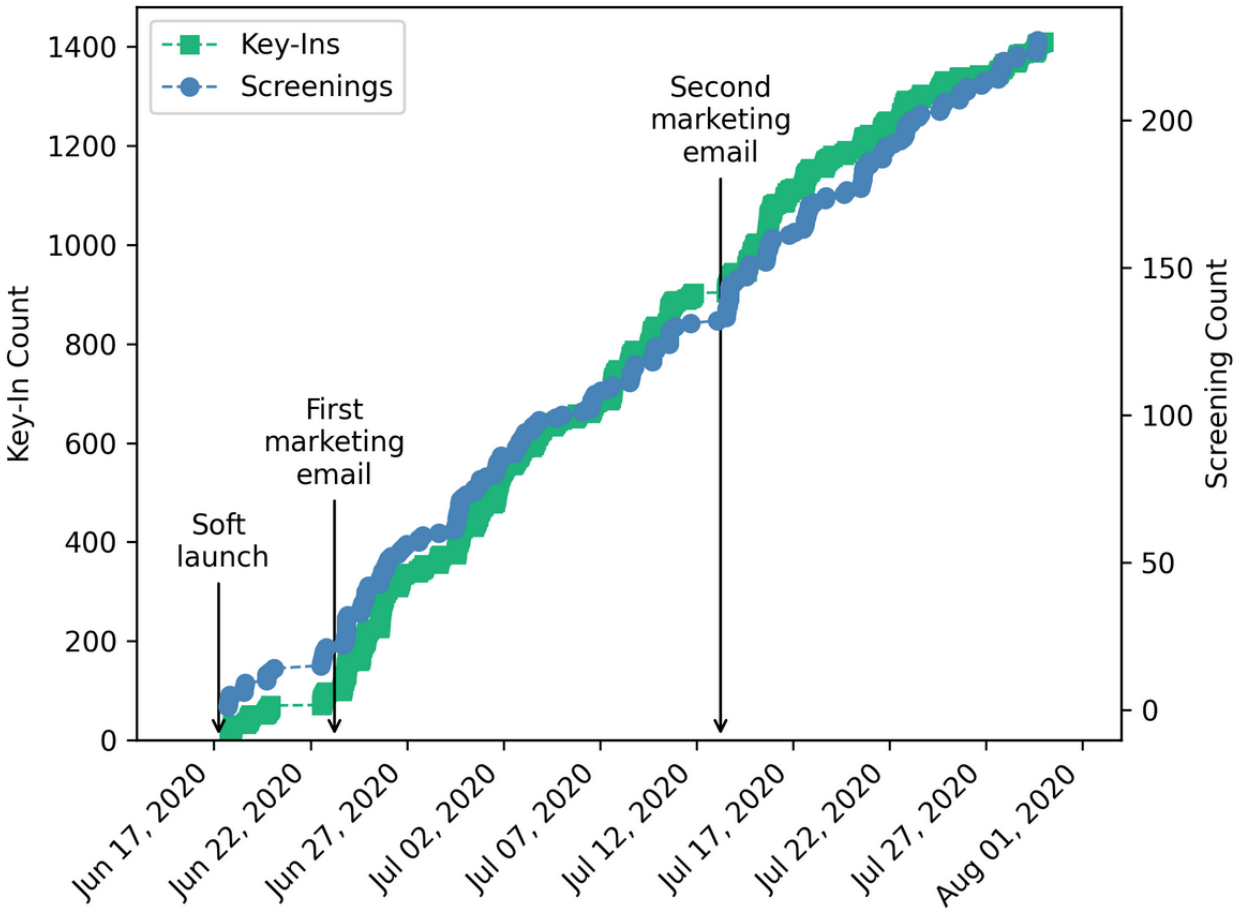
Results

Overall Usage

Over the 6-week pilot period, 45 participants created accounts. While our participants were not entirely from a single department, the majority were affiliated with the Department of Chemistry. For context, the Department of Chemistry has approximately 210 graduate students, postdoctoral fellows, faculty, and staff. During Phase 1 of the reopening, while operating at 33% capacity, 69 people were allowed to occupy space within the department; while at 50% capacity, this number increased to 105 people.

Of the 45 created accounts, 43 users logged in to the app at least once. These participants performed a total of 227 self-assessments and keyed in 1410 times at 48 distinct locations. Our soft launch period resulted in modest participant enrollments and app usage (Figure 7). On June 23, the first recruitment email was sent and the weekly raffle was instituted, and both participant sign-ups and app usage increased substantially. A second recruitment email was sent out approximately mid-way through the study (timed to avoid conflict with the July 4 holiday closure); however, it had little impact on app usage.

Figure 7. Usage of key-ins and screenings throughout the duration of the study along with key project events.



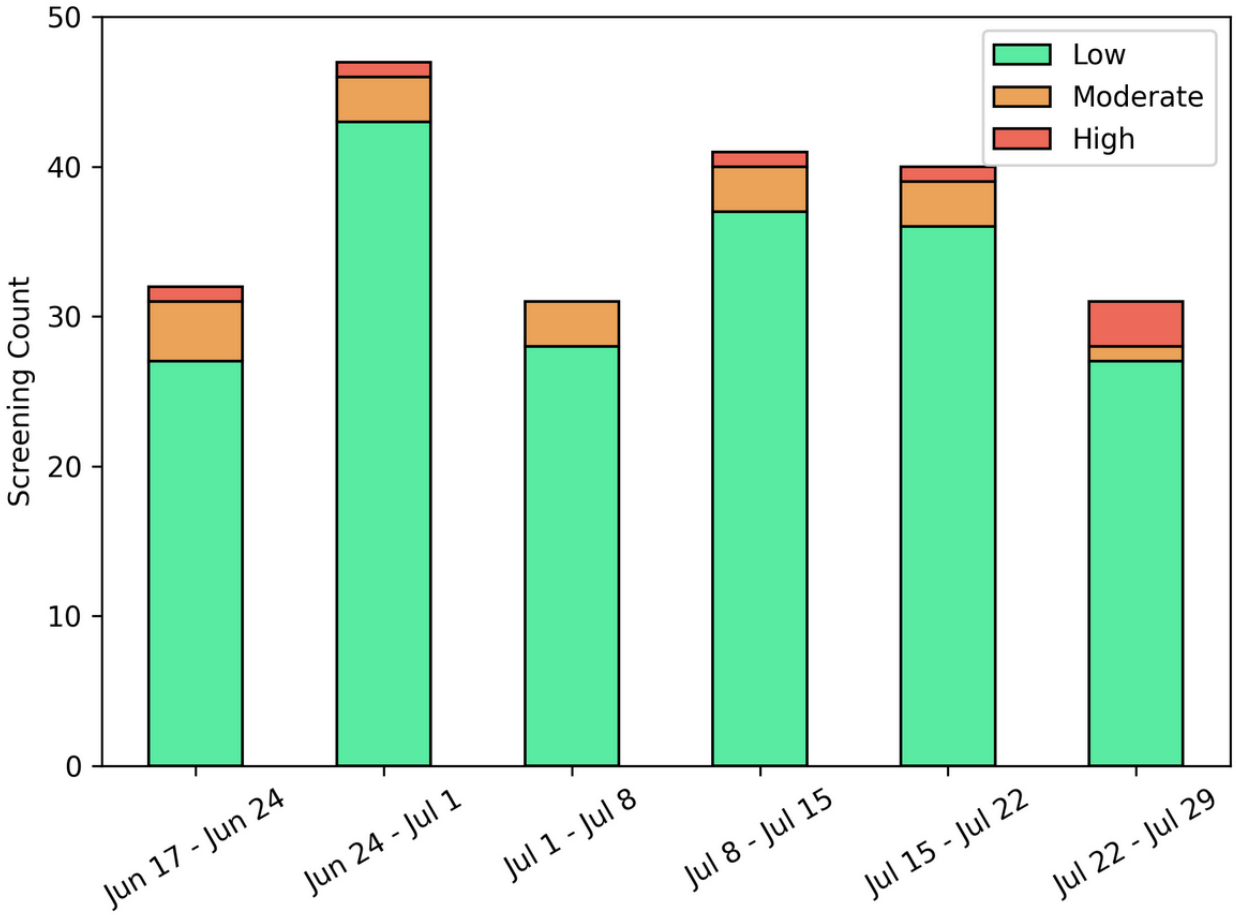
In the following sections, we analyze the self-assessments, key-ins, and contact tracing cases that resulted from this usage. Of the 45 individual users, only 26 completed the follow-up survey in its entirety, and 4 returned the survey incomplete (n=30, 66.6%). A total of 15 users did not complete the final follow-up survey. All of the users who completed the survey in some capacity provided demographic information including age, race, and gender.

Self-assessment and Key-In Usage

Self-assessments were performed by 89% (40/45) of users. The majority of the assessments (202/227, 89%) indicated low risk (ie, asymptomatic with no known exposures), 7.5% (17/227) of self-assessments were of moderate risk (ie, nonzero scores less than 3), and 3.5% (8/227) of self-assessments were of high risk (ie, scores of 3 or more) (Figure 8). Accounting for the different dates of user account creation, users performed 1.02

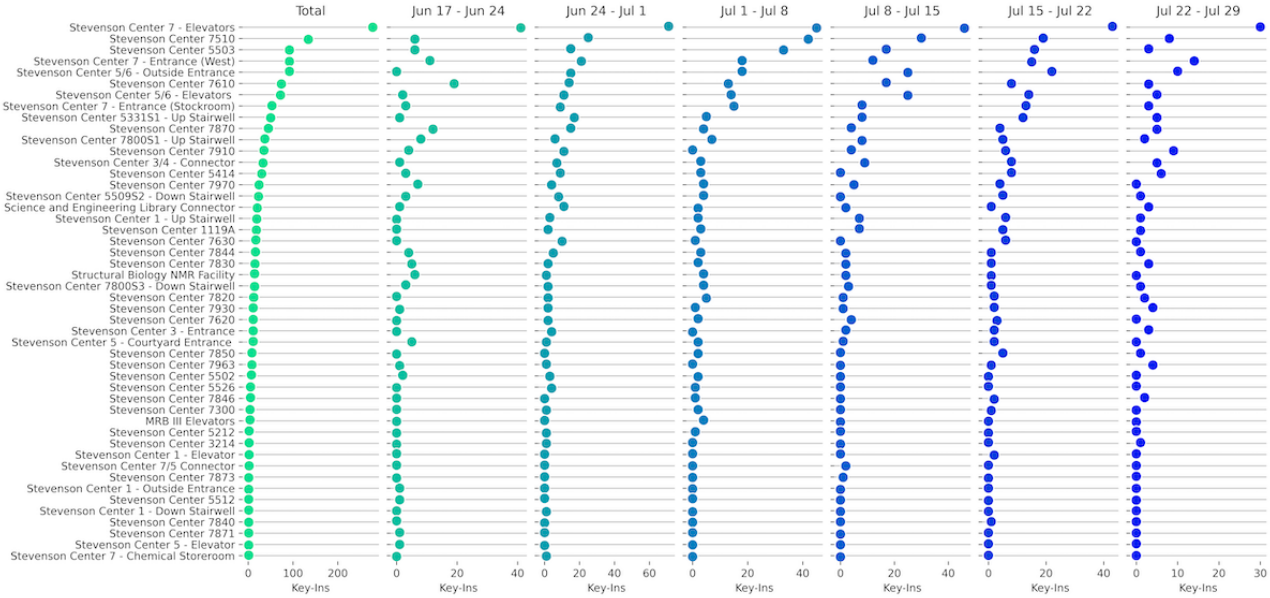
self-assessments per week (Multimedia Appendix 3). There were slight variations in the total number of screenings per week, with the fewest screenings being taken over the July 4 holiday week. The number of high-risk screenings increased in the final week as a result of a confirmed positive case within the study population.

Figure 8. Weekly counts of user self-assessments classified as low, moderate, or high risk.



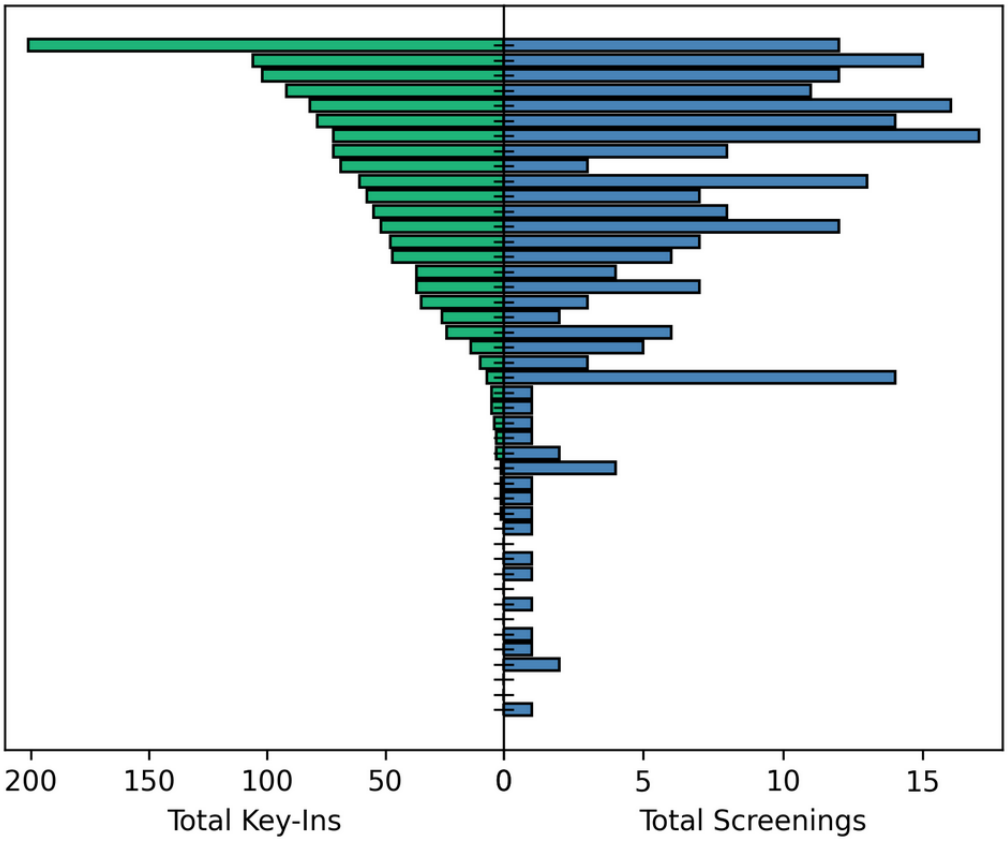
Key-ins were performed by 32 different users and occurred at 48 unique locations. Accounting for the variation in dates of user account creation, on average, users keyed in 6.75 times per week (Multimedia Appendix 4). Only 67% (n=48) of the 71 locations with flyers were actually used by the participants. The 5 most commonly visited locations accounted for almost 50% (688/1410) of all key-ins (Figure 9). Several of the most frequented locations are expected: the most central elevator at the heart of Stevenson Center Building 7 (the home building for the majority of our users) and multiple building entrances. While several locations saw a substantial increase or decrease in usage from week to week, possibly in part due to our enrollment size being small and our results therefore subject to the fluctuations of individual schedules, the rate of usage at the most frequented locations remained roughly constant from week to week.

Figure 9. Key-ins per location for each week.



While Figure 7 suggests a proportional relationship between the usage of the self-assessment tool and the key-in feature, app usage was not evenly distributed among our users, as would be expected with a new technology [42]. Figure 10 shows the total key-ins and screenings for our users (each user being a horizontal line on the y-axis), sorted by the number of key-ins for that user. The top of the graph shows that we had several high-volume participants who utilized both features of the app frequently. Conversely, there were 5 accounts that never keyed in or took a self-assessment (2 of which never logged in after creating an account). A total of 10 users did not use the app beyond their first self-assessment. Interestingly, several users appear to have used the self-assessment tool disproportionately compared to their use of the key-in feature. This is possibly tied to the increase in remote work for those individuals relative to their on-campus hours, or potential concerns over privacy after initial usage of the app.

Figure 10. A comparison of the total key-ins and screenings for each user in the pilot study. The total key-ins per user are shown on the left (green), while the number of screenings is displayed on the right-hand side (blue).



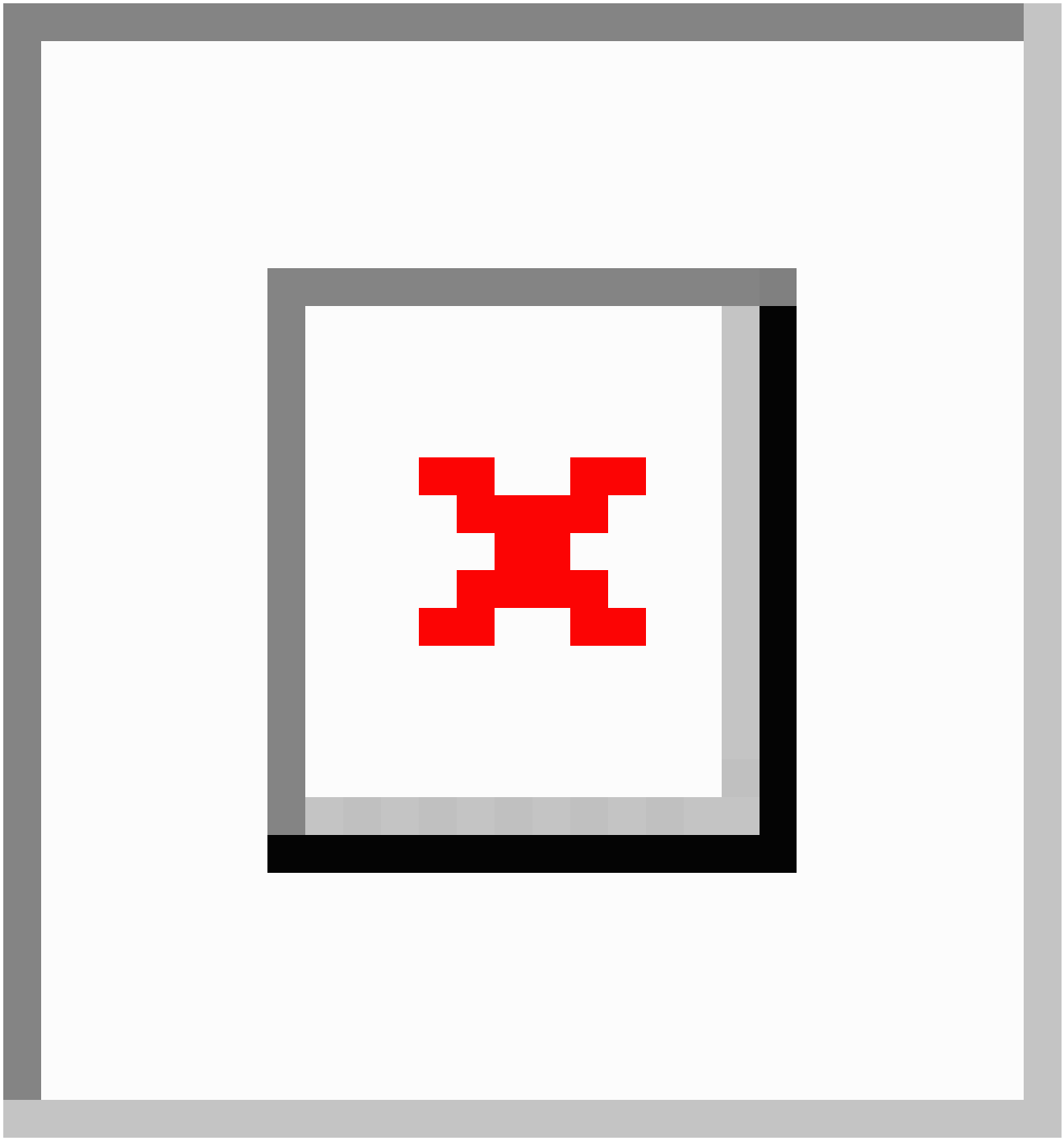
Contact Tracing

The potential for interactions, even in a small number of people, is large (Multimedia Appendix 5). Our app has two approaches for contact tracing using this individualized spatiotemporal data: manual and automatic. In manual contact tracing, administrators can search for a user by name or email address, find the locations that these users have visited, and identify any other users that keyed in to these locations within the overlap window (Multimedia Appendix 6). In automatic mode, a contact tracing case is created after each self-assessment that indicates either symptoms of or potential exposure to COVID-19. Every case consists of a POI (the user that took the self-assessment), locations of interest (locations that the POI keyed in to during their case window), and contacts of interest (other users that keyed in to locations of interest within a predefined "overlap" window). While the manual mode is designed to augment traditional contact tracing with digital data, automatic contact tracing can be used to streamline this process by compiling lists of contacts and locations, and potentially automating some tasks (notifications, cleaning schedules, etc).

Over the duration of the study, 25 self-assessments indicated either symptoms of or potential exposure to COVID-19. The 25 cases came from 8 unique users, and in 19 of the cases, the POI keyed in to a location on campus after their assessment indicated they were NOT CLEAR. In the event of an at-risk self-assessment, our app makes a prominent recommendation that users isolate and assists them to identify testing locations nearby (Figures 4B and 6D), but our pilot did not have the authority to keep users away from campus. For the purposes of this pilot study, we did not collect self-reported information from users on whether they were tested after receiving a NOT CLEAR status.

Of the 19 cases where the POI keyed in at least once on campus, there were 26 unique locations affected. The cases are summarized in a network chart (Figure 11) where each green square represents a location, blue circles represent users, and the red circle represents the POI. Lines connecting the POI and locations represent key-ins at those locations during the case window. Lines connecting other users and these locations represent key-ins during the overlap window. For brevity, we have not included any cases where POIs had multiple NOT CLEAR self-assessments within the same case window in Figure 11. Several cases had no overlapping users, while in others the density of connected locations at risk and contacts at risk was markedly increased.

Figure 11. A network connectivity diagram showing person-of-interest (POI) key-ins to locations of interest, as well as key-ins by other users at those same locations within the overlap window.



All digital contact tracing algorithms have parameters that must be explored in order to optimize accuracy. In our automated algorithm, the following parameters could be adjusted: the reverse case window period, the forward case window period, and the overlap window. We explored the sensitivity of our results to each of these parameters. While the total number of cases is fixed by the results of the users' self-assessments, as expected, the key-ins per POI, number of locations at risk, and number of contacts at risk all increase as these windows increase (Multimedia Appendices 7-9).

Discussion

Principal Findings

The COVID-19 pandemic has brought disease control strategies to the general public's attention. The need for robust contact tracing is broadly understood, particularly as states, and consequently, educational institutions, move through their phased reopening plans. While the need is agreed upon, reports of the lack of contact tracing infrastructure highlight the space where digital contact tracing tools can be useful. In this work, we describe a pilot study of MyCOVIDKey, a digital contact tracing app. The app consists of recurring self-assessments and user key-ins, whereby a user scans a unique barcode to indicate

their presence at a location. A 6-week pilot study took place within the Stevenson Center Science and Engineering Complex, on the Vanderbilt University campus in Nashville, TN. The pilot study was successful, and after app revisions based on user feedback (presented in detail in Scherr et al [39]), MyCOVIDKey will be ready for wider-scale deployment to campus and office settings. In this study, we found two clear purposes that could be addressed with digital interventions like MyCOVIDKey: (1) the identification of contacts of a POI who could have potentially been exposed and (2) the identification of locations that POIs visited that may be candidates for enhanced cleaning. Both are expected to remain key needs throughout the duration of the pandemic, even after the distribution of a vaccine.

While the postpilot survey results are analyzed separately in greater detail [39], it is worth noting that from this data it was clear that the majority of MyCOVIDKey users were young: 73.3% (22/30) of respondents were aged 20-30 years, while 20% (6/30) were aged 30-40 years and 6.66% (2/30) were 41 years of age or older. In addition, 77% (23/30) of our users were graduate students engaged in research. This cohort represents a biased group that is more likely to adopt newer technologies, confident in utilizing mobile phone apps, and interested in participating in the pandemic response. As such, our users may have different usage patterns, concerns, and preferences than a larger campus population, or even more so compared to a nonacademic audience. This selection bias was unavoidable considering the location and timing of the study, and its impact should be further studied on larger populations.

As we developed our app, we made several key decisions that should be further explored. Some implementations of COVID-19 self-assessments for return-to-work purposes do not allow users to access buildings or floors of their office space if they exhibit symptoms. This study took the alternative approach of allowing users to continue keying in with an at-risk self-assessment. This decision was made primarily for two reasons: (1) our pilot study did not have the authority to deny the participants entry into buildings or send them home from work, as those decisions were left to the reopening guidelines from the university; (2) we believed that there was the likelihood that users with at-risk self-assessments would continue to enter the building, regardless of their MyCOVIDKey status, and it was preferential to obtain data on their locations while at risk. This decision, albeit with a small sample size, was validated by the result that 19 of the 25 NOT CLEAR statuses still keyed in on campus, which indicates minimal behavioral change occurred, in this study, based on the user's status. Ideally, symptomatic individuals would follow the app's recommendations and isolate until they have either received a negative diagnostic test result or their window for transmission has lapsed. While this could have resulted from the perception of a lack of enforcement authority of the study, it could have also been explained by any diagnostic testing results that users may have obtained during the study. We are unable to draw conclusions on compliance since we did not actively seek input on diagnostic testing results after a NOT CLEAR status. This lack of diagnostic backing for self-reported symptoms may have introduced some amount of information bias due to the reliance on user memory and self-reporting. Regardless, this highlights a clear distinction between contact tracing software and a "passport" that allows entry if you meet checkpoint criteria. Given the level of asymptomatic spread of COVID-19, we believe that such passports are meaningful when tied to recent diagnostic testing—and considerably less useful with self-assessments alone. This distinction becomes even more critical when entrance to a location is tied to an incentive, for instance financial incentives at work or social or educational incentives on campuses.

In this study, we noted several parameters in our automatic contact tracing algorithm that must be tuned. Using the CDC's guidelines of 6 feet or less for 15 minutes or more to denote a "true" contact event, there will always be false positives and false negatives associated with digital contact tracing tools. False positives generated by digital contact tracing tools will increase the workload for manual contact tracers. For instance, increasing the overlap window or the case window parameters of our system will increase the number of locations and potential contacts that need to be traced. This could potentially become overwhelming for manual contact tracers in large organizations or in populations where there is a relatively high positivity rate. In contrast, false negatives from digital contact tracing tools will rely on manual efforts to correctly be identified, or risk not knowing forward disease transmission. We therefore recommend that the sensitivity and specificity with our system, and likely other digital contact tracing tools, be optimized depending on the population size, the local disease prevalence, and the level of automation allowed by contact tracing. One option that could be implemented in parallel to relieve burden on manual contact tracing efforts is to allow automated digital tools to only take action based on events that can be classified with a high degree of confidence. Based on the necessary tuning of parameters, it is our belief that digital contact tracing tools still serve best as a complement to manual contact tracing efforts, and not as a standalone replacement. This is not to minimize their importance. In fact, we believe they are an essential supplement to the realistic infrastructure constraints observed with manual contact tracing. When used appropriately, they can reduce the burden facing manual contact tracers by offloading certain inquiries and tasks.

While all contact tracing tools share the same goals, our technology has some notable differences from other approaches. MyCOVIDKey does not rely on Bluetooth or GPS to identify potential contact events; rather it relies on users to scan a barcode that identifies locations that they enter. This has technical advantages over the latter technologies, namely its ability to distinguish users in the same room from those separated by walls or even on different floors, as well as enhanced user privacy. Its primary disadvantages are that it does not capture potential exposures that could occur in transit between locations, and that it requires users to actively participate rather than rely on a continuous, automated data stream. While passive data collection is attractive to users due to the minimal effort required, it does come with increased privacy concerns—particularly as the sale of user location data for marketing purposes has become commonplace [43-47].

An additional limitation of our platform compared to others is the inability to determine how long users stayed at a particular

location or to determine their proximity to other users. Since users are only asked to key in upon entrance, and not exit, in the current version of the software, determining the overlap window's forward time limit is a challenge. Using the default overlap window of 30 minutes, our results for contacts of interest would count relatively harmless events like the keying in of 2 users to an elevator 25 minutes apart. However, it would miss events that may be noteworthy; for instance, key-ins to a classroom or laboratory that take place an hour apart, but where the POI has not yet left the room. A simple improvement is to allow organizations to define specific windows of interaction for different types of locations. This could more accurately reflect, for instance, that the timescales spent in elevators (seconds) is fundamentally different than time spent in classrooms (minutes) and in research labs (hours). An alternative approach to remedying this would be to ask users to key in at stations upon exiting as well. While this would place more burden on users and may therefore negatively affect continued usage outside of the consistent user group, it would provide the needed closure on user activity to ensure a more prescriptive assessment of risky interactions.

In this study, we did not ask users with a NOT CLEAR status if they received diagnostic testing to confirm or override their status in the app. The primary objectives of this study were to evaluate the usage of the platform and not to compare self-reported symptoms with diagnostic testing. Therefore, users who were identified as NOT CLEAR and considered a POI may have received a negative result from a SARS-CoV-2 diagnostic test and would be allowed to safely return to campus. While inclusion of this information has obvious utility, as in the aforementioned case, its implementation may be (depending on the disclosing party, any verification of the test results with the provider, and the user's parent organization that is utilizing this information) subject to regulation by the Health Insurance Portability and Accountability Act. In our postpilot survey, we did ask users about their experiences with COVID-19 testing. While explained in greater detail in our analysis of the postpilot survey data [39], one MyCOVIDKey user did report a positive diagnostic test for SARS-CoV-2. Importantly, MyCOVIDKey was able to accurately identify this person as a POI (this user was symptomatic, and their self-assessment indicated high risk), the locations they had visited in the buildings, and their contacts of interest. Given that the on-campus population was small due to local government safer-at-home orders and the university's emphasis on remote work where possible, the university's manual contact tracing team had sufficient capacity to handle this case. This reiterates that while the study setting was ideal for this pilot trial, MyCOVIDKey is perhaps most appropriate for settings where contact tracing infrastructure is not able to handle the volume of cases without additional support.

The usage of MyCOVIDKey during the pilot period closely followed the diffusion of innovation theory. The pilot had a group of early adopters that eagerly took on the platform. This core group was responsible for driving early usage and likely had a positive impact on encouraging new sign-ups and continued usage among their peers. Our pilot study launched without an organizational mandate or directive to use our app. In the absence of this, we made use of a weekly raffle to incentivize usage and participation. Businesses and higher-education institutes have the authority to give employees and students such an order. Forced mandates, however, could be met with resentment and resistance that would negatively affect their usage and undermine their objectives. So while it is understood that there is a critical threshold of users that must be reached in order for these tools to be effective [18], organizations must carefully balance the concerns of their members with public health needs when deciding how to meet this threshold.

Conclusion

Contact tracing is an essential component of any response to an epidemic, and digital contact tracing platforms are poised to play a large role in the current COVID-19 pandemic. In this paper, we have described one such tool, MyCOVIDKey, and a pilot evaluation of its usefulness in a university setting. We were able to identify several potential roles of digital contact tracing supplements, including the identification of potential contacts of at-risk individuals and resource allocation for local testing and building facilities management. While our platform, and these results, are directly applicable to campus communities, they are extensible to the reopening of businesses and communities at large as well. Although more studies are necessary to understand how variations in both the district and national levels could affect uptake in disparate populations and to develop effective mobile health implementation approaches [48], digital health interventions will likely be utilized worldwide. All organizations must make decisions on how best to integrate these tools into existing pandemic response infrastructure, as well as how to address potential concerns over data ownership and stewardship, while still reaching a critical threshold of necessary users for these tools to be effective. With a better understanding of the broader utility of MyCOVIDKey and apps like it, refinements will be made to simultaneously enhance the app's usability and security. Our pilot study shows that MyCOVIDKey can address the needs of many academic institutions and businesses as they begin to reopen.

Acknowledgments

This research was supported in part by the Amazon Web Services Diagnostic Development Initiative. The authors would like to acknowledge technical assistance from Scandit, as well as cloud services support from Jason Bradley at Vanderbilt University Information Technology.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The user privilege hierarchy for MyCOVIDKey, along with the data that each user class can access.
[PNG File , 523 KB - mhealth_v9i3e24275_app1.png]

Multimedia Appendix 2
The home screen of MyCOVIDKey displays information about the user's current MyCOVIDKey status, allows users to perform self-assessments, key in to new locations, and view simple usage statistics. Certain features are disabled, and the text is adjusted to reflect a user's current status: (A) NO STATUS for new accounts, and (B) EXPIRED status.
[PNG File , 466 KB - mhealth_v9i3e24275_app2.png]

Multimedia Appendix 3
An overview schematic of the MyCOVIDKey automated contact tracing algorithm.
[PNG File , 118 KB - mhealth_v9i3e24275_app3.png]

Multimedia Appendix 4
The probability density of key-ins and screenings per week.
[PNG File , 57 KB - mhealth_v9i3e24275_app4.png]

Multimedia Appendix 5
A network graph of all key-ins, regardless of time. Green squares represent locations, blue circles represent individual users, and lines represent key-ins from a user at a specific location.
[PNG File , 592 KB - mhealth_v9i3e24275_app5.png]

Multimedia Appendix 6
The manual contact tracing portal provides contact information for the POI, the locations that the user keyed in to during the search window, as well as overlapping users at those locations.
[PNG File , 125 KB - mhealth_v9i3e24275_app6.png]

Multimedia Appendix 7
A sensitivity analysis of how varying the reverse (left) and forward (right) case windows affects the amount of POIs that keyed-in during their case window.
[PNG File , 131 KB - mhealth_v9i3e24275_app7.png]

Multimedia Appendix 8
A sensitivity analysis of how varying the reverse (left) and forward (right) case windows affects how many key-ins had overlapping users.
[PNG File , 125 KB - mhealth_v9i3e24275_app8.png]

Multimedia Appendix 9
A sensitivity analysis of how varying the contact overlap window affects the number of overlapping users (contacts of interest).
[PNG File , 105 KB - mhealth_v9i3e24275_app9.png]

References

1. Coronavirus Disease 2019 (COVID-19). Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/php/contact-tracing/contact-tracing-plan/contact-tracing.html [accessed 2020-08-10]
2. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
3. Nussbaumer-Streit B, Mayr V, Dobrescu AI, Chapman A, Persad E, Klerings I, et al. Quarantine alone or in combination with other public health measures to control COVID-19: a rapid review. Cochrane Database Syst Rev 2020 Apr 08;4:CD013574. [doi: 10.1002/14651858.CD013574] [Medline: 32267544]
4. Engle S, Stromme J, Zhou A. Staying at Home: Mobility Effects of COVID-19. SSRN Journal. Preprint posted online Apr 15, 2020. [doi: 10.2139/ssrn.3565703]
5. Czeisler ME, Tynan MA, Howard ME, Honeycutt S, Fulmer EB, Kidder DP, et al. Public Attitudes, Behaviors, and Beliefs Related to COVID-19, Stay-at-Home Orders, Nonessential Business Closures, and Public Health Guidance - United States, New York City, and Los Angeles, May 5-12, 2020. MMWR Morb Mortal Wkly Rep 2020 Jun 19;69(24):751-758 [FREE Full text] [doi: 10.15585/mmwr.mm6924e1] [Medline: 32555138]

6. Gostin LO, Wiley LF. Governmental Public Health Powers During the COVID-19 Pandemic: Stay-at-home Orders, Business Closures, and Travel Restrictions. JAMA 2020 Jun 02;323(21):2137-2138. [doi: 10.1001/jama.2020.5460] [Medline: 32239184]
7. Nashville COVID-19 Response. URL: https://www.asafenashville.org/ [accessed 2020-08-11]
8. Davidson County COVID-19 Dashboard. Story Map Series. URL: https://nashville.maps.arcgis.com/apps/MapSeries/index.html?appid=30dd8aa876164e05ad6c0a1726fc77a4 [accessed 2020-08-10]
9. Return to Campus. Vanderbilt University. URL: https://www.vanderbilt.edu/coronavirus/ [accessed 2020-08-11]
10. Larochelle MR. "Is It Safe for Me to Go to Work?" Risk Stratification for Workers during the Covid-19 Pandemic. N Engl J Med 2020 Jul 30;383(5):e28. [doi: 10.1056/nejmp2013413]
11. Communities, Schools, Workplaces, & Events. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/coronavirus/2019-ncov/community/colleges-universities/ihe-testing.html [accessed 2020-08-10]
12. Shaw WS, Main CJ, Findley PA, Collie A, Kristman VL, Gross DP. Opening the Workplace After COVID-19: What Lessons Can be Learned from Return-to-Work Research? J Occup Rehabil 2020 Sep 19;30(3):299-302 [FREE Full text] [doi: 10.1007/s10926-020-09908-9] [Medline: 32562129]
13. Callahan A, Steinberg E, Fries JA, Gombar S, Patel B, Corbin CK, et al. Estimating the efficacy of symptom-based screening for COVID-19. NPJ Digit Med 2020 Jul 13;3(1):95 [FREE Full text] [doi: 10.1038/s41746-020-0300-0] [Medline: 32695885]
14. Gostic K, Gomez A, Mummah R, Kucharski A, Lloyd-Smith J. Estimated effectiveness of symptom and risk screening to prevent the spread of COVID-19. eLife 2020;9:e55570 [FREE Full text] [doi: 10.7554/eLife.55570v]
15. Jones J, Sullivan PS, Sanchez TH, Guest JL, Hall EW, Luisi N, et al. Similarities and Differences in COVID-19 Awareness, Concern, and Symptoms by Race and Ethnicity in the United States: Cross-Sectional Survey. J Med Internet Res 2020 Jul 10;22(7):e20001 [FREE Full text] [doi: 10.2196/20001] [Medline: 32614778]
16. Bai Y, Yao L, Wei T, Tian F, Jin D, Chen L, et al. Presumed Asymptomatic Carrier Transmission of COVID-19. JAMA 2020 Apr 14;323(14):1406-1407 [FREE Full text] [doi: 10.1001/jama.2020.2565] [Medline: 32083643]
17. Contact Tracing for COVID-19. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/php/contact-tracing/contact-tracing-plan/contact-tracing.html [accessed 2021-03-09]
18. Baraniuk C. Covid-19 contact tracing: a briefing. BMJ 2020 May 13;369:m1859. [doi: 10.1136/bmj.m1859] [Medline: 32404304]
19. Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med 2008 Mar 25;5(3):e74 [FREE Full text] [doi: 10.1371/journal.pmed.0050074] [Medline: 18366252]
20. Beidas RS, Buttenheim AM, Feuerstein-Simon R, Kilaru AS, Asch DA, Volpp KG, et al. Optimizing and Implementing Contact Tracing through Behavioral Economics. NEJM Catalyst. 2020 Jul 23. URL: https://catalyst.nejm.org/doi/abs/10.1056/CAT.20.0317 [accessed 2020-08-10]
21. Salathé M, Althaus CL, Neher R, Stringhini S, Hodcroft E, Fellay J, et al. COVID-19 epidemic in Switzerland: on the importance of testing, contact tracing and isolation. Swiss Med Wkly 2020 Mar 09;150:w20225 [FREE Full text] [doi: 10.4414/smw.2020.20225] [Medline: 32191813]
22. Sun K, Viboud C. Impact of contact tracing on SARS-CoV-2 transmission. The Lancet Infectious Diseases 2020 Aug;20(8):876-877 [FREE Full text] [doi: 10.1016/S1473-3099(20)30357-1]
23. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. The Lancet Infectious Diseases 2020 Aug;20(8):911-919 [FREE Full text] [doi: 10.1016/S1473-3099(20)30287-5] [Medline: 32353347]
24. Rothe C, Schunk M, Sothmann P, Bretzel G, Froeschl G, Wallrauch C, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N Engl J Med 2020 Mar 05;382(10):970-971 [FREE Full text] [doi: 10.1056/NEJMc2001468] [Medline: 32003551]
25. Wang S, Ding S, Xiong L. A New System for Surveillance and Digital Contact Tracing for COVID-19: Spatiotemporal Reporting Over Network and GPS. JMIR Mhealth Uhealth 2020 Jun 10;8(6):e19457 [FREE Full text] [doi: 10.2196/19457] [Medline: 32499212]
26. Yasaka TM, Lehrich BM, Sahyouni R. Peer-to-Peer Contact Tracing: Development of a Privacy-Preserving Smartphone App. JMIR Mhealth Uhealth 2020 Apr 07;8(4):e18936 [FREE Full text] [doi: 10.2196/18936] [Medline: 32240973]
27. Privacy-Preserving Contact Tracing - Apple and Google. Apple Inc. URL: https://www.apple.com/covid19/contacttracing [accessed 2020-08-10]
28. Digital Contact Tracing Tools. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/php/contact-tracing/contact-tracing-plan/digital-contact-tracing-tools.html [accessed 2021-03-09]
29. Vaudenay S. Report 2020/399: Analysis of DP3T. Cryptology ePrint Archive. 2020 Apr 8. URL: http://eprint.iacr.org/2020/399 [accessed 2021-03-09]
30. Azad M, Arshad J, Akmal A, Riaz F, Abdullah S, Imran M, et al. A First Look at Privacy Analysis of COVID-19 Contact Tracing Mobile Applications. arXiv. Preprint posted online Aug 16, 2020 [FREE Full text]
31. Gvili Y. Report 2020/428: Security Analysis of the COVID-19 Contact Tracing Specifications by Apple Inc. and Google Inc. Cryptology ePrint Archive. 2020 Apr 16. URL: http://eprint.iacr.org/2020/428 [accessed 2021-03-09]

32. Baumgärtner L, Dmitrienko A, Freisleben B, Gruler A, Höchst J, Kühlberg J, et al. arXiv. Preprint posted online Nov 6, 2020 [FREE Full text]
33. Bengio Y, Janda R, Yu YW, Ippolito D, Jarvie M, Pilat D, et al. The need for privacy with public digital contact tracing during the COVID-19 pandemic. The Lancet Digital Health 2020 Jul;2(7):e342-e344. [doi: 10.1016/s2589-7500(20)30133-3]
34. Lin L, Martin TW. How Coronavirus is eroding privacy. Wall Street Journal. 2020 Apr 15. URL: https://www.wsj.com/articles/coronavirus-paves-way-for-new-age-of-digital-surveillance-11586963028 [accessed 2021-08-10]
35. Servick K. Cellphone tracking could help stem the spread of coronavirus. Is privacy the price? Science. 2020 Mar 22. URL: https://www.sciencemag.org/news/2020/03/cellphone-tracking-could-help-stem-spread-coronavirus-privacy-price [accessed 2021-08-10]
36. Kahn J, Johns Hopkins Project on Ethics and Governance of Digital Contact Tracing Technologies. Digital Contact Tracing for Pandemic Response: Ethics and Governance Guidance. Baltimore: Johns Hopkins University Press; 2020.
37. Ethical considerations to guide the use of digital proximity tracking technologies for COVID-19 contact tracing. World Health Organization. 2020 May 28. URL: https://www.who.int/publications-detail-redirect/WHO-2019-nCoV-Ethics_Contact_tracing_apps-2020.1 [accessed 2021-08-10]
38. Kahn Gillmor D. Principles for technology-assisted contact-tracing. ACLU. 2020 Apr 16. URL: https://www.aclu.org/report/aclu-white-paper-principles-technology-assisted-contact-tracing [accessed 2021-08-10]
39. Scherr TF, DeSousa J, Moore C, Hardcastle A, Wright DW. App Usage and Usability Impressions of a Barcode-Based Digital Contact Tracing Platform for COVID-19: Survey Study. JMIR Public Health Surveill 2021 Mar;7(3):e25859 [FREE Full text] [doi: 10.2196/25859] [Medline: 33630745]
40. MyCOVIDKey. Vanderbilt University. 2020. URL: https://www.mycovidkey.com/ [accessed 2021-03-09]
41. Get Tested COVID-19. 2020. URL: https://get-tested-covid19.org/ [accessed 2021-03-09]
42. Rogers EM. Diffusion of Innovations, 4th Edition. New York, NY: The Free Press; 2010.
43. Riederer C, Erramilli V, Chaintreau A, Krishnamurthy B, Rodriguez P. For sale : your data: by : you. New York, NY: Association for Computing Machinery; 2011 Presented at: Proceedings of the 10th ACM Workshop on Hot Topics in Networks; November; Cambridge, MA p. 1-6. [doi: 10.1145/2070562.2070575]
44. Ashworth L, Free C. Marketing Dataveillance and Digital Privacy: Using Theories of Justice to Understand Consumers' Online Privacy Concerns. J Bus Ethics 2006 Aug 26;67(2):107-123. [doi: 10.1007/s10551-006-9007-7]
45. Martin KD, Murphy PE. The role of data privacy in marketing. J of the Acad Mark Sci 2017 Mar;45(2):135-155. [doi: 10.1007/s11747-016-0495-4]
46. Ghosh A, Roth A. Selling privacy at auction. New York, NY: Association for Computing Machinery; 2011 Presented at: Proceedings of the 12th ACM Conference on Electronic Commerce; August; Honolulu, HI p. 199-208. [doi: 10.1145/1993574.1993605]
47. Krishnamurthy B. I know what you will do next summer. SIGCOMM Comput Commun Rev 2010 Oct 22;40(5):65-70. [doi: 10.1145/1880153.1880164]
48. Scherr T, Moore C, Thuma P, Wright D. Evaluating Network Readiness for mHealth Interventions Using the Beacon Mobile Phone App: Application Development and Validation Study. JMIR Mhealth Uhealth 2020 Jul 28;8(7):e18413 [FREE Full text] [doi: 10.2196/18413] [Medline: 32720909]

Abbreviations

CDC: Centers for Disease Prevention and Control
POI: person of interest
QR: quick response code

Edited by L Buis; submitted 11.09.20; peer-reviewed by J Heo, J King; comments to author 13.10.20; revised version received 20.11.20; accepted 25.02.21; published 26.03.21.

Please cite as:
Scherr TF, Hardcastle AN, Moore CP, DeSousa JM, Wright DW
Understanding On-Campus Interactions With a Semiautomated, Barcode-Based Platform to Augment COVID-19 Contact Tracing: App Development and Usage
JMIR Mhealth Uhealth 2021;9(3):e24275
URL: https://mhealth.jmir.org/2021/3/e24275
doi:10.2196/24275
PMID:33690142

Original Paper

Analyzing the Essential Attributes of Nationally Issued COVID-19 Contact Tracing Apps: Open-Source Intelligence Approach and Content Analysis

Jan-Patrick Weiß[1], MSc; Moritz Esdar[1], MA; Ursula Hübner[1], PhD

Health Informatics Research Group, Faculty of Business Management and Social Sciences, University of Applied Sciences Osnabrueck, Osnabrück, Germany

Corresponding Author:
Ursula Hübner, PhD
Health Informatics Research Group
Faculty of Business Management and Social Sciences
University of Applied Sciences Osnabrueck
Caprivistraße 30a
Osnabrück,
Germany
Phone: 49 5419692012
Email: u.huebner@hs-osnabrueck.de

Abstract

Background: Contact tracing apps are potentially useful tools for supporting national COVID-19 containment strategies. Various national apps with different technical design features have been commissioned and issued by governments worldwide.

Objective: Our goal was to develop and propose an item set that was suitable for describing and monitoring nationally issued COVID-19 contact tracing apps. This item set could provide a framework for describing the key technical features of such apps and monitoring their use based on widely available information.

Methods: We used an open-source intelligence approach (OSINT) to access a multitude of publicly available sources and collect data and information regarding the development and use of contact tracing apps in different countries over several months (from June 2020 to January 2021). The collected documents were then iteratively analyzed via content analysis methods. During this process, an initial set of subject areas were refined into categories for evaluation (ie, coherent topics), which were then examined for individual features. These features were paraphrased as items in the form of questions and applied to information materials from a sample of countries (ie, Brazil, China, Finland, France, Germany, Italy, Singapore, South Korea, Spain, and the United Kingdom [England and Wales]). This sample was purposefully selected; our intention was to include the apps of different countries from around the world and to propose a valid item set that can be relatively easily applied by using an OSINT approach.

Results: Our OSINT approach and subsequent analysis of the collected documents resulted in the definition of the following five main categories and associated subcategories: (1) background information (open-source code, public information, and collaborators); (2) purpose and workflow (secondary data use and warning process design); (3) technical information (protocol, tracing technology, exposure notification system, and interoperability); (4) privacy protection (the entity of trust and anonymity); and (5) availability and use (release date and the number of downloads). Based on this structure, a set of items that constituted the evaluation framework were specified. The application of these items to the 10 selected countries revealed differences, especially with regard to the centralization of the entity of trust and the overall transparency of the apps' technical makeup.

Conclusions: We provide a set of criteria for monitoring and evaluating COVID-19 tracing apps that can be easily applied to publicly issued information. The application of these criteria might help governments to identify design features that promote the successful, widespread adoption of COVID-19 tracing apps among target populations and across national boundaries.

(JMIR Mhealth Uhealth 2021;9(3):e27232) doi:10.2196/27232

KEYWORDS

COVID-19; contact tracing; app; protocol; privacy; assessment; review; surveillance; monitoring; design; framework; feature; usage

Introduction

The COVID-19 pandemic has drawn unprecedented attention to public health measures and exposed weaknesses in governmental pandemic management efforts throughout all nations. In particular, the evidence on presymptomatic virus transmissions [1] and the large variance in severity between asymptomatic disease progression and deadly disease progression [2] has delineated the limits of centrally coordinated and executed test and trace programs. This has led to increased attention and the development of COVID-19 tracing apps for smartphones, which have been deemed as feasible and potentially effective tools for pandemic management [3]. As a result, apps are now seen as potentially important tools for supporting pandemic management; they provide a promising opportunity to complement the tracing efforts of local health authorities.

The effectiveness of contact tracing apps however depends on how many people in a given population use the app [4], which poses a unique sociotechnical challenge. A study by Trang et al [5] suggests that the large-scale willingness to use such apps is closely tied to the design factors of the app itself. This is especially true for the following design factors: privacy preservation, transparency, and convenience. Furthermore, when expanding the scope of app use to an international level, demands for transparent and open app developments become even more pressing. Contact tracing apps should not only have the ability to function in one country, but also have the ability to be interoperable with the solutions used in other countries (or the same app should be used in multiple countries), as SARS-CoV-2 spreads around the globe with little regard for national borders. At the same time, app users must be assured that a given contact tracing app is not misused as a surveillance tool by the issuing government. Such misuse might compromise human rights, privacy, and the acceptance of app use. Thus, technical app designs, especially those that relate to privacy-preserving elements, have far-reaching consequences.

The challenges of effectively using contact tracing apps are fundamentally similar across governments. This calls for the close monitoring of the different approaches to guiding governments in choosing the most promising apps. Although several studies have investigated the acceptance factors of COVID-19 tracing apps [5-8], limited research has been conducted on cross-country comparisons. O'Neill et al [9] published an overview of several national COVID-19 tracing apps. However, their study was not conducted in a scientific context, and they used a rather small set of evaluation criteria [9]. Ming et al [10] assessed various COVID-19–related apps across several countries but only briefly touched on these apps' tracing functionalities. Furthermore, checklists and item sets that pertain to the evaluation of general mobile health apps [11-15] do not seem particularly suited for evaluating contact tracing apps, as most criteria do not apply or are too unspecific. Therefore, our goal was to develop an item set that was suitable for describing and monitoring nationally coordinated COVID-19 contact tracing apps. This item set could provide a framework for describing the key technical features of such apps and monitoring their use based on widely available information.

Methods

In order to develop a suitable evaluation framework that accounts for feasibility in terms of the public availability of relevant information, we conducted an inductive procedure in which we screened the development and use of COVID-19 tracing apps in different countries over several months (from June 2020 to January 2021). The process of identifying and monitoring the apps was performed by using an open-source intelligence (OSINT) approach [16]. This procedure lends itself to the collection of data and information on new and emerging situations wherein there is only limited scientific knowledge available. Although this method involves drawing on scientific publications in journals, it goes beyond this and draws on a multitude of publicly available sources such as (1) grey literature (technical reports, preprints, white papers, and business documents); (2) government data (reports, press conferences, websites, and speeches); (3) conventional media (newspapers, magazines, radio, and television); and (4) the internet (web-based publications, blogs, and social media).

This approach enables the gathering of the most current information in a timely manner and helps to quickly assess the different cultural contexts of various countries. Textbox 1 provides an overview of the sources that were used in this study.

Two authors with backgrounds in medical informatics and information systems conducted parallel searches for publicly available information. Collected materials were reviewed by a third author who had a background in medical informatics. The third author also checked the collected information to determine whether the same information from different sources matched or contradicted each other. The search engines Google, DuckDuckGo, and Google Scholar were used for the search (search term combination format: "COVID-19 tracing app name" + "country name"). In addition, Wikipedia entries for COVID-19 contact tracing apps that were written in the native language of the apps' respective countries were translated by using services that were provided by Google and Microsoft (for information that was hardcoded in pictures or fliers, the Google Translate app was used). This was done to find references for COVID-19 contact tracing apps, as Wikipedia pages that were written in the native languages of the apps' respective countries were more up to date than those that were written in English and German (the two languages spoken by the authors). Our main information source was official information that was provided by the countries themselves (ie, technical reports from the apps' developers or media statements that were provided by official representatives of the countries). Media and academic publications were used to verify the information that was provided by the official sources of the countries or to find information that was not provided by official sources.

Textbox 1. Overview of the type of sources in this study.

Academic publications (9 documents) [5-8,17-21]

- Peer-reviewed information from conferences or journals
- Can also include nonpeer-reviewed sources like theses and dissertations

Public government data (28 documents) [22-48]

- Official sources, such as public government reports, press releases, government websites, or development repositories (eg, GitLab and GitHub)

Grey literature (5 documents) [49-53]

- Technical reports or preprints

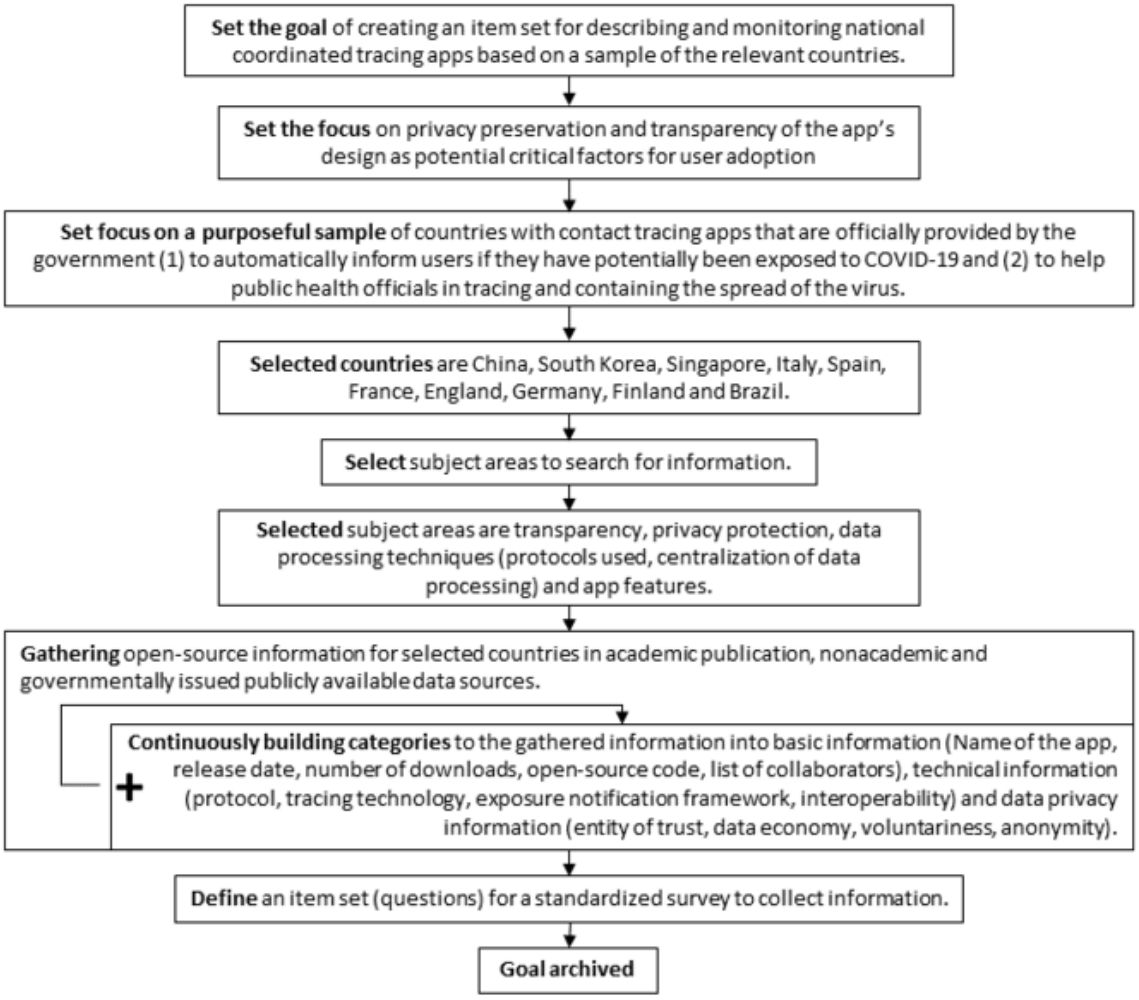
Media (11 documents) [9,54-63]

- News and articles from newspapers, magazines, radio, television, and podcasts (eg, British Broadcasting Corporation, The Guardian, and The New York Times), especially those on the internet
- News and articles from social media websites (eg, Twitter, YouTube, etc)

The initial entry points for researching relevant information revolved around subject areas that were reported to be important design factors in previous studies [5,18-21,53] on tracing apps, such as transparency, privacy protection, data processing techniques (the protocols used and the centralization of data processing), and app features. The documents and websites that were gathered from selected countries were then analyzed by using content analysis methods, which were applied throughout several iterations to inductively extract coherent, topical categories that emerged from the available information [64]. Each category was refined into subcategories and a set of related features. These were subsequently paraphrased as question items. The final item set was then reapplied to the selected countries.

The iteration process, its findings, and the final item set framework are presented in the Results section. Figure 1 summarizes our entire methodological approach. These methodological steps were applied to information from a sample of countries with contact tracing apps that were officially provided by the government to (1) automatically inform users about whether they have potentially been exposed to COVID-19 and (2) help public health officials with tracing and containing the spread of SARS-COV-2. The sample of national tracing apps was purposefully selected; our intention was to include apps from different countries from around the world. By looking at various countries, we aimed to explore the variety and variability in the technical attributes of COVID-19 tracing apps. The nationally commissioned apps from the following countries (in alphabetical order) constituted the purposeful sample: Brazil, China, Finland, France, Germany, Italy, Singapore, South Korea, Spain, and the United Kingdom (England and Wales). These countries reflect different regions that had different experiences with the COVID-19 outbreak and the containment of the pandemic. We also expected different technological approaches across the selected countries.

Figure 1. Sequence of methodological steps.



Results

Summary of Categories

After the initial screening of the sources, several clusters of information needs emerged. These were refined and summarized into five categories (Textbox 2). These categories were then used to derive the final item set.

In the following sections, the information we collected for each category is presented. These data are based on our review of the following apps: Corona Warn App (Germany), TousAntiCovid (France), Immuni (Italy), Radar COVID (Spain), Koronavilkku (Finland), NHS (National Health Service) COVID-19 (England and Wales), TraceTogether (Singapore), Self-Diagnosis app (South Korea), Self-Quarantine app (South Korea), and Coronavirus-SUS (Sistema Único de Saúde; Brazil). Of note, China has no single app and instead uses various other apps (eg, WeChat) that have integrated tracing solutions. The order of this list mirrors the availability of information (most to least available) in the 10 selected countries.

Textbox 2. Categories and related information needs.

Background information

- Open-source code
- How transparent is the development process?
- Is information regarding the development process and the app disseminated? If yes, how is it done?

Purpose and workflow

- Is the exact position of users being tracked?
- Can health authorities contact users via the app?
- Is additional data gathered (eg, for epidemiological analysis)?

Technical information

- Which protocol is implemented?
- Which tracing technology is used?
- Is the Google/Apple Exposure Notification System used?
- Is it interoperable with other apps?

Privacy protection

- Is a centralized or decentralized approach used?
- Can the gathered data be deleted?
- Does the user have to provide any personal information?
- Is the use of the app mandatory (eg, access controls)?

Availability and use

- When was the app released?
- How often was the app downloaded?

Background Information

After investigating the first category (ie, background information), we found that the provision of primary background information by the issuers of the apps (Table 1) was essential for assessing the issuers' intentions for the app and the apps' implications for data privacy. The apps of Germany [41], France [24], Italy [38], Spain [65], Finland [66], England and Wales [67], and Singapore [23] have open-source repositories of their code bases and websites [32,34-37,42,68,69], which act as dedicated information hubs for citizens and tourists. Except for Singapore's and South Korea's apps, all countries' apps were commissioned by their respective health and digital ministries. However, the apps were mainly developed, deployed, and maintained by the private industry sector. Singapore and South Korea have dedicated ministries that are capable of developing contact tracing apps with minor support from the private industry sector. Only the German development team from SAP SE (Systeme, Andwendungen, Produkte in der Datenverarbeitung, Societas Europaea) provided some form of insight into their app development process via a wide range of documents, including the scoping document, technical documentation, and an issue tracker that can be found in their publicly accessible GitHub repository [41]. There was no documentation found for the apps of Brazil, China, and South Korea (eg, documents with information about open-source code or dedicated websites). As such, various publicly available sources, like statements from press conferences, government-issued guidelines, coverage by traditional media, and social media, had to be searched to obtain information about these countries' apps [28,48,55,57,60,70-73].

Table 1. Background information provided by the issuers of the apps.

App	Country	Open-source repository	Information websites
Corona Warn App	Germany	Yes	Yes
TousAntiCovid	France	Yes	Yes
Immuni	Italy	Yes	Yes
Radar COVID	Spain	Yes	Yes
Koronavilkku	Finland	Yes	Yes
NHS[a] COVID-19	England and Wales	Yes	Yes
TraceTogether	Singapore	Yes	Yes
Self-Diagnosis app	South Korea	No	No
Self-Quarantine app	South Korea	No	No
Coronavirus-SUS[b]	Brazil	No	No
Various integrated tracing functions in apps (eg, WeChat)	China	No	No

[a]NHS: National Health Service.

[b]SUS: Sistema Único de Saúde.

Purpose and Workflow

Our analysis of the second category (ie, the purpose of the app) showed that all of the apps that were included in this study were developed to help public health officials with tracing and containing the spread of SARS-CoV-2, as stated in the materials that were publicly accessible at the time of our search [26-28,31,32,34-37,42,48,57,60,65,68,72,74,75]. Except for the apps of China and South Korea, all apps notify their users about whether they came into contact with another user who tested positive for COVID-19. China's app informs its users via a 3-color code system (green, yellow, or red) and uses a quick response (QR) code that can be scanned by authorities and businesses. Green indicates a person who has unrestricted movement, yellow indicates a person who must undergo a 7-day home quarantine, and red indicates a patient with confirmed COVID-19 who needs to be quarantined for 14 days. In this system, it is not clear as to how these codes are assigned and removed [55,60,71]. South Korea uses a slightly different approach compared to those of the other countries. Most apps automatically match possible COVID-19 contact events. However, South Korea allows its users to self-report their health status with the Self-Diagnosis app. If the reported symptoms are indicative of a SARS-CoV-2 infection, health authorities can intervene. If a person tests positive for COVID-19, a case officer from the local government is assigned, and the person who tested positive must report their symptoms via another app called the "Self-Quarantine app." If people with COVID-19 leave their designated quarantine areas, they and the case officers receive an alert [57].

Technical Information

With regard to the third category (ie, technical information), the Decentralized Privacy-Preserving Proximity Tracing (DP-3T) [49], Temporary Contact Numbers (TCN) [52], BlueTrace [22] and Pan-European Privacy-Preserving Proximity Tracing (PEPP-PT) [76] protocols were found to be used as reference implementations for the various apps (Table 2). DP-3T is used by almost all European countries except for France, which uses PEPP-PT with their own reference implementation for the Robust and Privacy-Preserving Proximity Tracing (ROBERT) protocol [51]. The protocols used for the apps of China and South Korea were not disclosed. The apps of Brazil, Finland, and England and Wales use the Google/Apple Exposure Notification System (GAEN) but do not provide any specific information about which reference protocol they used for app development. The DP-3T, TCN, BlueTrace, and ROBERT protocols all use Bluetooth-based proximity detection approaches. These Bluetooth-based proximity detection approaches use different mathematical models to calculate the likelihood of a potential exposure. The key parameters for these models are the time of being near other devices, the distance to other devices, and the results of an assessment for the contagiousness of a person with COVID-19. The sensitivity of these parameters can be adjusted by the issuer of the app, depending on the number of infections. South Korea relies on users' self-reports, and when users are quarantined, the country relies on their GPS to alert health authorities about users leaving their quarantine zones [72]. China uses QR codes (to be able to access certain locations, these QR codes have to be scanned upon entry) and other undisclosed sensor data and metadata from smartphones [60,71,73]. Germany, Italy, Brazil, Spain, and England and Wales use the GAEN. The European Union has provided interoperability specifications for cross-border transmission chains between the approved apps that European Union states are planning to implement [29]. In October 19, 2020, the apps of Germany, Ireland, and Italy became interoperable with each other. This was made possible by the European interoperability gateway service, which will include other countries (eg, Finland) in the future [30,31]. Singapore and South Korea have dedicated ministries that are capable of developing contact tracing apps with only minor support from the private industry sector.

Table 2. Technical information of the apps.

App	Country	Protocol	Tracing technology	GAEN[a]	Interoperable with apps from other countries
Corona Warn App	Germany	DP-3T[b] and TCN[c]	Bluetooth	Yes	Yes
TousAntiCovid	France	ROBERT[d]	Bluetooth	No	No
Immuni	Italy	DP-3T	Bluetooth	Yes	Yes
Radar COVID	Spain	DP-3T	Bluetooth	Yes	Yes
Koronavilkku	Finland	—[e]	Bluetooth	Yes	Yes
NHS[f] COVID-19	England and Wales	—	Bluetooth	Yes	Yes
TraceTogether	Singapore	OpenTrace	Bluetooth	No	No
Self-Diagnosis app	South Korea	—	Reported data	No	No
Self-Quarantine app	South Korea	—	Reported data and GPS	No	No
Coronavirus-SUS[g]	Brazil	—	Bluetooth	Yes	No
Various integrated tracing functions in apps (eg, WeChat)	China	—	Quick response code, sensor data, and metadata	No	No

[a]GAEN: Google/Apple Exposure Notification System.

[b]DP-3T: Decentralized Privacy-Preserving Proximity Tracing.

[c]TCN: Temporary Contact Numbers.

[d]ROBERT: Robust and Privacy-Preserving Proximity Tracing.

[e]Not available.

[f]NHS: National Health Service.

[g]SUS: Sistema Único de Saúde.

Privacy Protection

The different protocols have various implications with regard to the privacy of users (the fourth category). China and South Korea use fully centralized protocols and have no methods in place for anonymizing collected data (Table 3). The other apps from Germany, France, Italy, Spain, England and Wales, Singapore, and Brazil use different protocols but function similarly to each other. All countries' apps use Bluetooth to perform tracing; they do not use any other methods of geolocation (eg, using GPS data). As such, the apps only record how long two devices (smartphones) come into contact but not where contacts occur. Singapore's app requires mobile phone numbers to be registered. However, aside from this feature, none of the apps collect any user-identifying data and only exchange alternating IDs via Bluetooth. If the GAEN is used, the tracing and identification of contact events with people who test positive for COVID-19 are locally conducted on users' smartphones in a decentralized manner. The apps of Singapore and France also trace contacts locally on smartphones via user logs. However, user logs must be uploaded to the centralized servers of health authorities for report processing. These servers identify contact events with people who test positive for COVID-19 and send warning messages to users. Additionally, Singapore has combined the function of its contact tracing app, TraceTogether, with the function of the SafeEntry check-in system. In this system, a QR code that contains users' contact information must be scanned before entering a location. This information is then uploaded to a government server. England and Wales have also implemented a similar venue check-in function directly into their app.

In the decentralized approach of Germany, Italy, Spain, Finland, England and Wales, and Brazil, user logs never leave the smartphone. However, if app users from these countries test positive for COVID-19, they can enable their apps to upload a key to the central server of health authorities. This key is then sent to all devices, and health authorities can subsequently derive device IDs and check if they match one of the encounters in the user log. The use of contact tracing apps in Germany, France, Italy, Spain, Finland, England and Wales, Brazil, and Singapore is completely voluntary. Furthermore, the apps, along with their data, can be removed and deleted at any time.

Table 3. Privacy protection.

App	Country	Contact tracing approach	Report processing approach	User can opt out	Anonymity (no user registration)
Corona Warn App	Germany	Decentralized	Decentralized	Yes	Yes
TousAntiCovid	France	Decentralized	Centralized	Yes	Yes
Immuni	Italy	Decentralized	Decentralized	Yes	Yes
Radar COVID	Spain	Decentralized	Decentralized	Yes	—[a]
Koronavilkku	Finland	Decentralized	Decentralized	Yes	Yes
NHS[b] COVID-19	England and Wales	Decentralized	Decentralized (tracing)/centralized (check-ins)	Yes	Yes
TraceTogether	Singapore	Decentralized	Centralized	Yes	Phone number
Self-Diagnosis app	South Korea	—	Centralized	No	Name, address, and phone number
Self-Quarantine app	South Korea	—	Centralized	No	Name, address, and phone number
Coronavirus-SUS[c]	Brazil	Decentralized	Decentralized	Yes	Yes
Various integrated tracing functions in apps (eg, WeChat)	China	Centralized	Centralized	No	No

[a]Not available.

[b]NHS: National Health Service.

[c]SUS: Sistema Único de Saúde.

Availability and Use

With regard to the fifth category (ie, availability and use), the use of contact tracing apps first occurred in Asian countries, starting with China [70] (integrated tracing functions in existing apps like WeChat or Alipay) and South Korea [28] (Self-Diagnosis app and Self-Quarantine app) in February 2020 and Singapore [77] (TraceTogether) in March 2020. Italy [74] (Immuni), France [75] (their app originally launched as StopCovid and relaunched as TousAntiCovid in October 2020 [78]), and Germany [27] (Corona Warn App) all released their contact tracing apps in June 2020. Brazil expanded the function of its app [48] (Coronavirus-SUS) so that citizens could seek information about COVID-19 (eg, noticeable fake news and recent outbreak locations) and report possible COVID-19 symptoms to check for a potential SARS-CoV-2 infection; a tracing function was added in August 2020. Most of the autonomous communities in Spain adopted the Radar COVID app after COVID-19 testing ended in other regions at the end of August 2020 [79]. Finland also released its app at the end of August 2020 [33]. After a failed attempt at developing a contact tracing app (a result of privacy issues), England and Wales abandoned its first app development approach and released a new app in September 2020. This app also used the GAEN [25,26,54].

Although most countries in our sample regularly reported download numbers, Brazil did not provide regular reports on recent download numbers (Table 4). There are also no comparable download numbers available for China and South Korea, as China integrated their tracing functions into popular existing apps and South Korea used a different tracing system in which not all citizens continuously use one app. For South Korea however, download numbers are available from the Google Play Store; Apple does not report on the number of downloads.

Table 4. Availability and use.

App	Country	Release date	Downloads, n (% of population)	Date reported	Source type
Corona Warn App	Germany	June 16, 2020	25.4 million (30.6%)	February 11, 2021	Official [40]
TousAntiCovid	France	June 2, 2020	11 million (16.4%)	December 8, 2020	Official [43]
Immuni	Italy	June 15, 2020	10.3 million (17.06%)	February 22, 2021	Official [44]
Radar COVID	Spain	August 2020	7.03 million (17%)	February 14, 2021	Official [45]
Koronavilkku	Finland	August 31, 2020	>2.5 million (45.3%)	November 5, 2021	Official [46]
NHS[a] COVID-19	England and Wales	September 24, 2020	21.8 million (36.7%)	February 10, 2021	Official [47]
TraceTogether	Singapore	March 20, 2020	4.2 million (73.7%)	February 23, 2021	Official [34]
Self-Diagnosis app	South Korea	March 2020	>500,000 (0.96%)	February 23, 2021	Google Play Store [61]
Self-Quarantine app	South Korea	March 2020	>500,000 (0.96%)	February 23, 2021	Google Play Store [62]
Coronavirus-SUS[b]	Brazil	August 1, 2020	10 million (4.77%)	December 8, 2020	Official [48] and press [63]
Various integrated tracing functions in apps (eg, WeChat)	China	February 2020	—[c]	—	—

[a]NHS: National Health Service.

[b]SUS: Sistema Único de Saúde.

[c]Not available.

Evaluation Framework

Conducting research on the collected materials and the five categories allowed us to create more specific subcategories and derive a potential item set that was suitable for describing and monitoring nationally coordinated tracing apps (Table 5). As more information was found through the OSINT-based search, the items were phrased in a more detailed manner. This proposed item set for a monitoring framework was applied to the apps of the 10 countries that were analyzed in this study (Multimedia Appendix 1). In total, 19 items were applied to the 11 apps of the 10 countries. This yielded 209 cells of information in the matrix. However, data for 22 cells were not publicly available. These 22 cells were mainly related to information from China, but they were also somewhat related to information from South Korea, Brazil, Spain, and England and Wales. The items that were missing information were primarily related to the protocols that countries used and data privacy issues, particularly those concerning conformance with data protection regulations and the secondary use of data (other purposes).

Table 5. Proposed item set for a monitoring framework.

Category, subcategory, and item number	Item
Background information	
Name	
Q1.1	What is the name of the app?
Open-source code	
Q1.2	Is there a publicly accessible repository of the source code?
Q1.3	Is the published source code up to date?
Public information	
Q1.4	Is there some form of material (eg, website, guideline, or Frequently Asked Questions page) for informing the public about the app?
Collaborators	
Q1.5	Which institutions worked together to develop, host, and maintain the app?
Purpose and workflow	
Warning process design	
Q2.1	What is the main mode of action for warning app users?
Secondary data use	
Q2.2	What other purposes are the data used for?
Technical information	
Protocol	
Q3.1	Which tracing protocols (eg, DP-3T[a], TCN[b], ROBERT[c], and BlueTrace) are implemented in the app?
Tracing technology	
Q3.2	Which tracing technology (eg, Bluetooth, GPS, and barcodes) is used by the app?
Exposure notification system	
Q3.3	Is the Google/Apple Exposure Notification System used?
Interoperability	
Q3.4	Is the app actively interoperable with apps from other countries?
Privacy protection	
Entity of trust	
Q4.1	Is the process of contact tracing (eg, tracking each contact event) centralized or decentralized? Provide a short description of the workflow.
Q4.2	Is the report processing approach (eg, matching contact events and informing the user) centralized or decentralized? Provide a short description of the workflow.
Q4.3	Are app data processed as mandated by data protection regulations?
Q4.4	Is the data automatically destroyed after a fixed period of time?
Q4.5	Can the user opt out?
Anonymity	
Q4.6	Does the user have to register any information (eg, mobile phone number, name, address, or date of birth)?
Availability and use	
Release date	
Q5.1	On which date could the app be downloaded by the public?
Number of downloads	
Q5.2	Are there officially reported download numbers?

[a]DP-3T: Decentralized Privacy-Preserving Proximity Tracing.

[b]TCN: Temporary Contact Numbers.

[c]ROBERT: Robust and Privacy-Preserving Proximity Tracing.

Discussion

Principal Results

COVID-19 tracing apps have become instruments in the arsenal of measures for fighting the spread of the disease. Many countries have adopted these instruments in their national pandemic management plans. This study therefore aimed to explore and identify the core features of COVID-19 tracing apps in order to provide a generic framework for conducting cross-country comparisons and monitoring apps over time. We found and screened publicly available information via an OSINT approach for analyzing selected subject areas, which served as initial entry points for collecting data. We then defined five categories to structure the evaluation framework. These categories were based on recurring information, and information for the framework was selected based on obtaining enough publicly available data. The categories were as follows: background information, purpose and workflow, technical information, privacy protection, and availability and use. Based on these categories, we constructed a set of specific items that could be used to evaluate the core features of COVID-19 tracing apps. In order to showcase the item set's usefulness, it was applied to 10 countries that nationally commissioned COVID-19 tracing apps (ie, Brazil, China, France, Germany, Italy, Spain, Finland, Singapore, South Korea, and the United Kingdom [England and Wales]). Our comparison revealed differences among each countries' apps, especially with regard to the centralization of the entity of trust and the overall transparency of the apps' technical makeup. The proposed item set will help researchers evaluate the spread and use of contact tracing apps within and across countries in the future.

Key Characteristics for Evaluating COVID-19 Tracing Apps

The application of OSINT in the emerging field of COVID-19 tracing apps heavily depends on the public availability of essential and comprehensive background information. Ideally, such information can be retrieved from an official website that informs the public and provides technical background information. Information on app development is even more useful, as open-source projects provide detailed insights and involve unrestricted code audits. In open-source projects, the source code and the entire project description are often transparently accessible. Detailed insights on app protocols and their implementation are interesting from both a technical and economic point of view (eg, increasing public acceptance); transparency likely facilitates trust, which is associated with increased public acceptance [6,7]. Our analysis revealed varying degrees of transparency across countries, as some countries disclosed most of the abovementioned information (Germany, France, Italy, Spain, the United Kingdom, and Singapore), while others provided noticeably less information (Brazil, China and South Korea).

As the motivation to download and use an app is the ultimate predictor of an app's success, information about download numbers, acceptance, and use are the most crucial for differentiating between successful and unsuccessful app designs. According to our initial findings, relative app download numbers were highest in Singapore, Finland, the United Kingdom, and Germany. Although download numbers are indicators of motivation, they do not necessarily reveal information about an app's actual use [17]. App use can only be measured centrally by certain methods. For instance, Germany's COVID-19 tracing app analyzes the total number of shared negative and positive test results [50]. However, the characteristics described in our framework could be used as possible predictors of app adoption rates in cross-country adoption research.

An app's technical design is essential not only for assessing its usefulness, but also for preserving privacy. In particular, the decision to use Bluetooth, GPS, or other means (eg, QR code scanning with corresponding protocols) determines whether the app serves the purpose of tracing or location tracking. The latter could potentially result in greater privacy concerns. Thus, tracking is unlikely to be accepted in many countries. On the other hand, several COVID-19 tracing app users have expressed their interest in obtaining more detailed information about close encounters with people with COVID-19 after receiving rather superficial information from their Bluetooth-based tracing apps [59]. This shows that COVID-19 tracing app developers face the challenge of striking the right balance between maximizing privacy preservation and maximizing usefulness (ie, the range of functionalities). It has yet to be determined which composition of app traits works best in which cultures. The use of Bluetooth is also closely related to the use of the GAEN. The GEAN system seems to be indispensable, as tracing apps that do not use it cannot run in the background of Apple/Android phones. Furthermore, many countries have shifted toward using the GAEN [58]. However, in the future, it might be desirable to find technical solutions for overcoming public health agencies' and governments' dependency on Google and Apple.

Another important technical feature in our item set is interoperability, which refers to an app's ability to operate in a synchronized manner with the COVID-19 tracing apps of other countries. This has become increasingly important, given the likely resurgence of international travel when the pandemic starts to recede. Du and colleagues [80] have stated that the risk of creating a useless Tower of Babel of contact tracing apps is very real, as contact tracing apps' inability to work across different countries renders them ineffective. Researchers should therefore identify which apps allow for cross-country use. This is particularly interesting to people who live in regions that neighbor other countries, which is quite common in Europe.

The acceptance of COVID-19 tracing apps hinges on both its perceived benefits and perceived barriers (eg, privacy concerns) [8]. The basic and technical information that are outlined in this study lay the foundation for privacy preservation. However, such information must be supplemented with data on a range of additional factors that relate to the entity of trust and anonymity. With regard to the entity of trust, we included indicators that can be used to determine whether a central or decentral data processing method is used. We found that this distinction could and should be determined for contact tracing

approaches and report processing approaches. Centrality is a particularly critical factor, as the providers of COVID-19 tracing apps are often governmental agencies, and many users might be reluctant to entrust their government with the exclusive handling of health-related data. In our sample, South Korea and China stood out because they opted for approaches with higher degrees of centrality. Furthermore, the aspects of centrality and a range of other well-known criteria for privacy preservation need to be considered. Therefore, we included automatic data deletion after a fixed period of time, the possibility to opt out, and the need to refrain from storing additional personal data (eg, mobile phone numbers) into our item set.

Limitations

Several limitations apply to our study. First and foremost, the OSINT approach depends on the public availability of reliable and trustworthy data and information. Second, the great majority of items could be answered for most, but not all, countries. This could have been due to missing information or language issues. Incorporating the insights of experts from the studied countries could have provided us with additional information and a more nuanced picture of app use/development. However, the approach we chose is more feasible, and its pragmatic nature allows for the flexible incorporation of various information sources in a quickly evolving field. Third, we developed the item set by analyzing a limited number of countries, and we cannot exclude the possibility that additional criteria would have emerged if we included additional countries in the analysis.

Conclusion and Outlook

Ours is one of the first studies to provide a set of criteria for evaluating nationally commissioned COVID-19 tracing apps and to apply such criteria to 10 industrialized countries. Although cross-country comparisons have been previously conducted [6,9], our study provides a more comprehensive yet relatively easy-to-apply evaluation framework that uses various technical factors from publicly available sources as potential determinants of app adoption.

The evaluation of COVID-19 tracing apps is crucial for the assessment of factors that facilitate these apps' widespread acceptance and usefulness within and across countries. The more people who accept and use the app, the better the virus can be contained. This has been demonstrated by simulation studies [4]. Given that transparency and privacy protection are crucial for building people's trust in apps, the technical features that are proposed in our framework might play an important role in promoting widespread app adoption. The initial application of our framework to the 10 countries in this study revealed differences in countries' app adoption rates. Based on the app download numbers, Finland, Germany, and the United Kingdom were more successful in deploying their apps because they chose transparent and decentralized contact tracing and report processing approaches. However, Singapore was the only city-state in our sample, and it had the highest app adoption rates.

The associations between technical features and app success must be addressed more thoroughly in future research. This can be done by analyzing a greater number of countries, and our framework provides the groundwork to do so. We made the item set very concise so that it can be easily shared with scientists in other countries via a collaborative approach or used to survey people in other countries via a crowdsourcing platform that is similar to that of Trang et al [5]. At the end of the pandemic, it will be interesting to see the role that apps actually played in the fight against the pandemic across different countries (ie, outside of those included in this study). Our framework will help studies that aim to evaluate apps' contributions to the overall management of pandemics, by providing core descriptors of COVID-19 warning apps that can be used as predictors.

Insights on the success factors of tracing apps might prove useful for designing general national health apps. Over time, tracing apps might be combined with health record apps, as there have already been discussions and instances of developers expanding their apps' features to include contact and symptom diaries, vaccination certificates, and educational resources for public health messaging.

Acknowledgments

This study was funded by the State of Lower Saxony, Germany (grant number: ZN 3062).

Authors' Contributions

UH initiated the study. JPW specified the theoretical framework, was supported by UH and ME, and conceptualized the research design. JPW collected and prepared the data with the support of ME, under the supervision of UH. All authors prepared the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Applied monitoring framework.
[XLSX File (Microsoft Excel File), 20 KB - mhealth_v9i3e27232_app1.xlsx]

References

1. Furukawa NW, Brooks JT, Sobel J. Evidence supporting transmission of severe acute respiratory syndrome coronavirus 2 while presymptomatic or asymptomatic. Emerg Infect Dis 2020 Jul;26(7):e201595 [FREE Full text] [doi: 10.3201/eid2607.201595] [Medline: 32364890]
2. Cen Y, Chen X, Shen Y, Zhang XH, Lei Y, Xu C, et al. Risk factors for disease progression in patients with mild to moderate coronavirus disease 2019-a multi-centre observational study. Clin Microbiol Infect 2020 Sep;26(9):1242-1247 [FREE Full text] [doi: 10.1016/j.cmi.2020.05.041] [Medline: 32526275]
3. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 2020 May 08;368(6491):eabb6936 [FREE Full text] [doi: 10.1126/science.abb6936] [Medline: 32234805]
4. Hinch R, Probert W, Nurtay A, Kendall M, Wymant C, Hall M, et al. Effective configurations of a digital contact tracing app: A report to NHSX. en. GitHub. 2020. URL: https://tinyurl.com/h6ctnxtx [accessed 2021-03-22]
5. Trang S, Trenz M, Weiger WH, Tarafdar M, Cheung CMK. One app to trace them all? Examining app specifications for mass acceptance of contact-tracing apps. Eur J Inf Syst 2020 Jul 27;29(4):415-428 [FREE Full text] [doi: 10.1080/0960085x.2020.1784046]
6. Altmann S, Milsom L, Zillessen H, Blasone R, Gerdon F, Bach R, et al. Acceptability of app-based contact tracing for COVID-19: Cross-country survey study. JMIR Mhealth Uhealth 2020 Aug 28;8(8):e19857 [FREE Full text] [doi: 10.2196/19857] [Medline: 32759102]
7. von Wyl V, Höglinger M, Sieber C, Kaufmann M, Moser A, Serra-Burriel M, et al. Drivers of acceptance of COVID-19 proximity tracing apps in Switzerland: panel survey analysis. JMIR Public Health Surveill 2021 Jan 06;7(1):e25701 [FREE Full text] [doi: 10.2196/25701] [Medline: 33326411]
8. Walrave M, Waeterloos C, Ponnet K. Adoption of a contact tracing app for containing COVID-19: A health belief model approach. JMIR Public Health Surveill 2020 Sep 01;6(3):e20572 [FREE Full text] [doi: 10.2196/20572] [Medline: 32755882]
9. O'Neill PH, Ryan-Mosley T, Johnson B. A flood of coronavirus apps are tracking us. Now it's time to keep track of them. MIT Technology Review. 2020 May 07. URL: https://www.technologyreview.com/2020/05/07/1000961/launching-mittr-covid-tracing-tracker/ [accessed 2021-02-12]
10. 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]
11. Levine DM, Co Z, Newmark LP, Groisser AR, Holmgren AJ, Haas JS, et al. Design and testing of a mobile health application rating tool. NPJ Digit Med 2020 May 21;3:74 [FREE Full text] [doi: 10.1038/s41746-020-0268-9] [Medline: 32509971]
12. Nouri R, Kalhori SRN, 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 [FREE Full text] [doi: 10.1093/jamia/ocy050] [Medline: 29788283]
13. Llorens-Vernet P, Miró J. Standards for mobile health-related apps: Systematic review and development of a guide. JMIR Mhealth Uhealth 2020 Mar 03;8(3):e13057 [FREE Full text] [doi: 10.2196/13057] [Medline: 32130169]
14. Khowaja K, Al-Thani D. New checklist for the heuristic evaluation of mHealth apps (HE4EH): Development and usability study. JMIR Mhealth Uhealth 2020 Oct 28;8(10):e20353 [FREE Full text] [doi: 10.2196/20353] [Medline: 33112252]
15. van Haasteren A, Vayena E, Powell J. The mobile health app trustworthiness checklist: Usability assessment. JMIR Mhealth Uhealth 2020 Jul 21;8(7):e16844 [FREE Full text] [doi: 10.2196/16844] [Medline: 32706733]
16. Lee S, Shon T. Open source intelligence base cyber threat inspection framework for critical infrastructures. In: 2016 Future Technologies Conference (FTC). 2016 Presented at: 2016 Future Technologies Conference (FTC); December 6-7, 2016; San Francisco, CA, USA p. 1030-1033. [doi: 10.1109/ftc.2016.7821730]
17. Riillo CAF, Peroni C, Sarracino F. Determinants of acceptability of contact tracing apps for COVID-19: Initial results from Luxembourg. Économie et statistiques: Working papers du STATEC. 2020 Dec. URL: https://statistiques.public.lu/catalogue-publications/economie-statistiques/2020/117-2020.pdf [accessed 2021-03-22]
18. Baumgärtner L, Dmitrienko A, Freisleben B, Gruler A, Höchst J, Kühlberg J, et al. Mind the GAP: Security & privacy risks of contact tracing apps. In: 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom). 2020 Presented at: 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom); December 29, 2020, to January 1, 2021; Guangzhou, China.
19. von Wyl V, Bonhoeffer S, Bugnion E, Puhan MA, Salathé M, Stadler T, et al. A research agenda for digital proximity tracing apps. Swiss Med Wkly 2020 Jul 13;150:w20324 [FREE Full text] [doi: 10.4414/smw.2020.20324] [Medline: 32672340]
20. Li J, Guo X. Global deployment mappings and challenges of contact-tracing apps for COVID-19. SSRN. Preprint posted online on May 24, 2020 [FREE Full text] [doi: 10.2139/ssrn.3609516]
21. Ahmed N, Michelin RA, Xue W, Ruj S, Malaney R, Kanhere SS, et al. A Survey of COVID-19 Contact Tracing Apps. IEEE Access 2020 Jul 20;8:134577-134601 [FREE Full text] [doi: 10.1109/access.2020.3010226]
22. BlueTrace Protocol. BlueTrace. URL: https://bluetrace.io/ [accessed 2020-11-09]
23. OpenTrace. GitHub. URL: https://github.com/OpenTrace-Community [accessed 2020-11-05]

24. TousAntiCovid sources. GitLab. URL: https://gitlab.inria.fr/stopcovid19 [accessed 2020-11-05]
25. Next phase of NHS coronavirus (COVID-19) app announced 2020. Department of Health and Social Care. 2020 Jun 18. URL: https://www.gov.uk/government/news/next-phase-of-nhs-coronavirus-covid-19-app-announced [accessed 2020-08-18]
26. NHS COVID-19 app launches across England and Wales: NHS COVID-19 app launches to help control coronavirus (COVID-19) transmission. Department of Health and Social Care. 2020 Sep 24. URL: https://www.gov.uk/government/news/nhs-covid-19-app-launches-across-england-and-wales [accessed 2020-10-20]
27. Veröffentlichung der Corona-Warn-App 2020. Die Bundesregierung. 2020 Jun 16. URL: https://www.bundesregierung.de/breg-de/themen/coronavirus/veroeffentlichung-der-corona-warn-app-1760892 [accessed 2020-08-17]
28. (Important) Installation of COVID-19 Self-diagnosis Mobile App when you visit Korea 2020. Embassy of the Republic of Korea in the Republic of Kenya. 2020 Mar 18. URL: http://overseas.mofa.go.kr/ke-en/brd/m_10538/view.do?seq=761314 [accessed 2020-08-17]
29. Corona-Warn-Apps sollen grenzüberschreitend funktionieren. Europäische Kommission. 2020 Jun 16. URL: https://ec.europa.eu/germany/news/20200616-corona-warn-apps_de [accessed 2020-08-18]
30. Mobile contact tracing apps in EU Member States. European Commission. URL: https://ec.europa.eu/info/live-work-travel-eu/coronavirus-response/travel-during-coronavirus-pandemic/mobile-contact-tracing-apps-eu-member-states_en [accessed 2021-01-11]
31. Coronavirus: EU interoperability gateway goes live, first contact tracing and warning apps linked to the system. European Commission. 2020 Oct 19. URL: https://ec.europa.eu/commission/presscorner/detail/en/ip_20_1904 [accessed 2020-10-21]
32. Koronavilkku. Finnish Institute for Health and Welfare. URL: https://koronavilkku.fi/en/ [accessed 2021-02-11]
33. Koronavilkku has been downloaded a million times already!. Finnish Institute for Health and Welfare. 2020 Sep 01. URL: https://thl.fi/en/web/thlfi-en/-/koronavilkku-has-been-downloaded-a-million-times-already- [accessed 2021-02-11]
34. TraceTogether. Government of Singapore. URL: https://www.tracetogether.gov.sg/ [accessed 2020-11-05]
35. SafeEntry - National digital check-in system. Government of Singapore. URL: https://www.safeentry.gov.sg/ [accessed 2020-11-05]
36. Radar covid19. Government of Spain. URL: https://radarcovid.gob.es/ [accessed 2020-11-05]
37. Immuni - Sito Ufficiale. Presidenza del Consiglio dei Ministri. URL: https://www.immuni.italia.it/ [accessed 2020-11-05]
38. Immuni - Commissario straordinario per l'emergenza Covid-19. GitHub. URL: https://github.com/immuni-app [accessed 2020-11-05]
39. Infektionsketten digital unterbrechen mit der Corona-Warn-App: Die Corona-Warn-App ist ein wichtiger Baustein der Pandemiebekämpfung. Robert Koch Institut. URL: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/WarnApp/Warn_App.html [accessed 2020-08-18]
40. Übersicht zu aktuellen und früherer Zahlen und Fakten zur Corona-Warn-App. Robert Koch Institute. URL: https://tinyurl.com/rcwz74f4 [accessed 2021-02-11]
41. Corona-Warn-App. GitHub. URL: https://github.com/corona-warn-app [accessed 2020-11-05]
42. Corona-Warn-App Open Source Project. The authors of the Corona-Warn-App open-source project. URL: https://www.coronawarn.app/en/ [accessed 2020-11-05]
43. Application TousAntiCovid millions de téléchargements. République Française. 2020 Dec 08. URL: https://www.vie-publique.fr/en-bref/277126-tousanticovid-11-millions-de-telechargements [accessed 2021-02-24]
44. Immuni - I numeri di Immuni. Presidenza del Consiglio dei Ministri. URL: https://www.immuni.italia.it/dashboard.html [accessed 2021-02-24]
45. Descargas Radar COVID. Gobierno de España. URL: https://radarcovid.gob.es/en/estadisticas/descargas-radar [accessed 2021-02-24]
46. Finnish IFH. Koronavilkku has been downloaded more than 2.5 million times – widespread use increases the app's effectiveness. Finnish Institute for Health and Welfare. 2020 Nov 05. URL: https://thl.fi/en/web/thlfi-en/-/koronavilkku-has-been-downloaded-more-than-2.5-million-times-widespread-use-increases-the-app-s-effectiveness [accessed 2021-02-24]
47. What is the latest data about NHS COVID-19 app downloads? National Health Service England. URL: https://faq.covid19.nhs.uk/article/KA-01368 [accessed 2021-02-24]
48. Penido A. Aplicativo Coronavírus-SUS vai alertar contatos próximos de pacientes com COVID-19. Fiocruz. 2020 Aug 11. URL: https://www.bio.fiocruz.br/index.php/br/noticias/1923-aplicativo-coronavirus-sus-vai-alertar-contatos-proximos-de-pacientes-com-covid-19 [accessed 2020-08-18]
49. DP^3T: Decentralized Privacy-Preserving Proximity Tracing. GitHub. URL: https://github.com/DP-3T [accessed 2020-11-09]
50. Diagnosis key analysis (dka). GitHub. URL: https://github.com/micb25/dka/blob/master/README.en.md [accessed 2021-01-08]
51. ROBERT -- ROBust and privacy-presERving proximity Tracing protocol. GitHub. URL: https://github.com/ROBERT-proximity-tracing [accessed 2020-11-09]
52. TCN Coalition and LFPH have merged. Linux Foundation Public Health. URL: https://tcn-coalition.org/ [accessed 2020-11-09]

53. Cho H, Ippolito D, Yu YW. Contact tracing mobile apps for COVID-19: Privacy considerations and related trade-offs. ArXiv. Preprint posted online on March 30, 2020 [FREE Full text]
54. Cellan-Jones R, Kleinman Z. NHS Covid-19 app: 12m downloads - and lots of questions. BBC News Service. 2020 Sep 28. URL: https://www.bbc.com/news/technology-54326267 [accessed 2020-10-20]
55. Davidson H. China's coronavirus health code apps raise concerns over privacy. The Guardian. 2020 Apr 01. URL: https://www.theguardian.com/world/2020/apr/01/chinas-coronavirus-health-code-apps-raise-concerns-over-privacy [accessed 2020-08-17]
56. Dillet R. French contact-tracing app StopCovid has been activated 1.8 million times but only sent 14 notifications. TechCrunch. 2020 Jun 23. URL: https://tinyurl.com/4jfv7hne [accessed 2020-08-18]
57. Kim M. South Korea is watching quarantined citizens with a smartphone app. MIT Technology Review. 2020 Mar 06. URL: https://www.technologyreview.com/2020/03/06/905459/coronavirus-south-korea-smartphone-app-quarantine/ [accessed 2020-08-17]
58. Sato M. Why some countries suspended, replaced, or relaunched their covid apps. MIT Technology Review. 2020 Dec 23. URL: https://www.technologyreview.com/2020/12/23/1015557/covid-apps-contact-tracing-suspended-replaced-or-relaunched/ [accessed 2021-01-08]
59. Welche dieser Probleme sind Ihnen bereits bei der Nutzung der Corona-Warn-App begegnet? Wirtschaftswoche. URL: https://civey.com/umfragen/11258/welche-dieser-probleme-sind-ihnen-bereits-bei-der-nutzung-der-corona-warn-app-begegnet [accessed 2021-01-08]
60. Xinhua NA. I've seen Hangzhou at 4 a.m., just for the early "no code" - "health code" behind the return to work. Xinhua News Agency. 2021 Feb 20. URL: https://baijiahao.baidu.com/s?id=1659147516916399925 [accessed 2020-08-17]
61. Quarantine-Report • Self-Check. Google Play. URL: https://play.google.com/store/apps/details?id=com.mohw.corona [accessed 2021-02-24]
62. 자가격리자 안전보호. Google Play. URL: https://play.google.com/store/apps/details?id=kr.go.safekorea.sqsm [accessed 2021-02-24]
63. Butcher I. App Coronavírus-SUS supera 10 milhões de downloads e 1,8 mil registros de exames positivos. 2020 Dec 08. URL: https://www.mobiletime.com.br/noticias/08/12/2020/coronavirus-sus-app-tem-10-milhoes-de-downloads-e-18-mil-registros-de-exames-positivos-para-covid-19/ [accessed 2021-02-24]
64. Krippendorff K. Content Analysis: An Introduction to Its Methodology Fourth Edition. Los Angeles, California: SAGE Publishing; 2018.
65. Radar COVID. GitHub. URL: https://github.com/RadarCOVID [accessed 2020-11-05]
66. Finnish Institute for Health and Welfare (THL). GitHub. URL: https://github.com/THLfi [accessed 2021-02-11]
67. NHSX. GitHub. URL: https://github.com/nhsx [accessed 2020-11-05]
68. TousAntiCovid. Ministère de l'Économie des Finances et de la Relance. 2020 Oct 22. URL: https://www.economie.gouv.fr/stopcovid [accessed 2020-11-05]
69. The NHS COVID-19 app support website. National Health Service England. URL: https://www.covid19.nhs.uk/ [accessed 2020-11-05]
70. China launches coronavirus 'close contact detector' app. BBC News Service. 2020 Feb 11. URL: https://www.bbc.com/news/technology-51439401 [accessed 2020-08-17]
71. Mehta I. China’s coronavirus detection app is reportedly sharing citizen data with police. The Next Web. 2020. URL: https://tinyurl.com/3teb7w6c [accessed 2020-08-17]
72. Self-isolated safety protection app wonder tapa Q&A 7. Ministry of Public Administration and Security. 2020 Mar 05. URL: https://tinyurl.com/y2sxtdyz [accessed 2020-08-18]
73. Mozur P, Zhong R, Krolik A. In coronavirus fight, China gives citizens a color code, with red flags. The New York Times. 2020 Mar 01. URL: https://www.nytimes.com/2020/03/01/business/china-coronavirus-surveillance.html [accessed 2020-08-18]
74. Immuni contact-tracing App: available on all devices. Ministry of Health. URL: https://tinyurl.com/2kctsrx9 [accessed 2020-08-17]
75. StopCovid: the project. Ministry of the Economy and Finance. 2020. URL: https://web.archive.org/web/20200805134409/https://www.economie.gouv.fr/stopcovid-le-projet [accessed 2020-08-17]
76. Pan-European Privacy-Preserving Proximity Tracing. PEPP-PT Initiative. URL: https://web.archive.org/web/20200731012945/https://www.pepp-pt.org/ [accessed 2020-08-04]
77. Bay J. TraceTogether - One month on. Medium. 2020 Apr 20. URL: https://blog.gds-gov.tech/one-month-on-6c882875b9ba [accessed 2020-08-17]
78. Info Coronavirus COVID-19 - Application Tousanticovid. Gouvernement.fr. URL: https://www.gouvernement.fr/info-coronavirus/tousanticovid [accessed 2021-02-24]
79. Mitchell D. Radar COVID contact tracing app is now active throughout Madrid. EuroWeekly. 2020 Oct 08. URL: https://www.euroweeklynews.com/2020/10/08/radar-covid-contact-tracing-app-is-now-active-throughout-madrid/ [accessed 2020-10-20]

80. Du L, Raposo VL, Wang M. COVID-19 contact tracing apps: A technologic Tower of Babel and the gap for international pandemic control. JMIR Mhealth Uhealth 2020 Nov 27;8(11):e23194 [FREE Full text] [doi: 10.2196/23194] [Medline: 33156804]

Abbreviations

DP-3T: Decentralized Privacy-Preserving Proximity Tracing
GAEN: Google/Apple Exposure Notification System
NHS: National Health Service
OSINT: open-source intelligence approach
PEPP-PT: Pan-European Privacy-Preserving Proximity Tracing
QR: quick response
SAP SE: Systeme, Andwendungen, Produkte in der Datenverarbeitung, Societas Europaea
SUS: Sistema Único de Saúde
TCN: Temporary Contact Numbers

Edited by L Buis; submitted 18.01.21; peer-reviewed by T Winkler, M Pobiruchin, T Heponiemi; comments to author 10.02.21; revised version received 03.03.21; accepted 12.03.21; published 26.03.21.

Please cite as:
Weiß JP, Esdar M, Hübner U
Analyzing the Essential Attributes of Nationally Issued COVID-19 Contact Tracing Apps: Open-Source Intelligence Approach and Content Analysis
JMIR Mhealth Uhealth 2021;9(3):e27232
URL: https://mhealth.jmir.org/2021/3/e27232
doi:10.2196/27232
PMID:33724920

Original Paper

Attitudes of General Practitioners Toward Prescription of Mobile Health Apps: Qualitative Study

Aline Sarradon-Eck[1,2], MD, PhD; Tiphanie Bouchez[3], MD; Lola Auroy[4], MA; Matthieu Schuers[5,6,7], MD, PhD; David Darmon[1,3], MD, PhD

[1]Aix Marseille University, INSERM, IRD, SESSTIM, Marseille, France

[2]Institut Paoli-Calmettes, CanBios UMR1252, Marseille, France

[3]Université Côte d'Azur, Rétines, Healthy, DERMG, Nice, France

[4]Université Grenoble Alpes, Centre National de la Recherche Scientifique, Sciences Po Grenoble, Pacte, Grenoble, France

[5]Department of General Medicine, Rouen University Hospital, Rouen, France

[6]Department of Biomedical Informatics, Rouen University Hospital, Rouen, France

[7]INSERM, U1142, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé (LIMICS), Sorbonne Université, Paris, France

Corresponding Author:
Aline Sarradon-Eck, MD, PhD
Aix Marseille University, INSERM, IRD, SESSTIM
UMR1252-CanBios-SESSTIM
232 Bb Ste Marguerite
Marseille, 13009
France
Phone: 33 4 91 22 33 09
Email: aline.sarradon@inserm.fr

Abstract

Background: Mobile health (mHealth) apps are a potential means of empowering patients, especially in the case of multimorbidity, which complicates patients' care needs. Previous studies have shown that general practitioners (GPs) have both expectations and concerns regarding patients' use of mHealth apps that could impact their willingness to recommend the apps to patients.

Objective: The aim of this qualitative study is to investigate French GPs' attitudes toward the prescription of mHealth apps or devices aimed toward patients by analyzing GPs' perceptions and expectations of mHealth technologies.

Methods: A total of 36 GPs were interviewed individually (n=20) or in a discussion group (n=16). All participants were in private practice. A qualitative analysis of each interview and focus group was conducted using grounded theory analysis.

Results: Considering the value assigned to mHealth apps by participants and their willingness or resistance to prescribe them, 3 groups were defined based on the attitudes or positions adopted by GPs: *digital engagement* (favorable attitude; mHealth apps are perceived as additional resources and complementary tools that facilitate the medical work, the follow-up care, and the monitoring of patients; and apps increase patients' compliance and empowerment); *patient protection* (related to the management of patient care and fear of risks for patients, concerns about patient data privacy and security, doubt about the usefulness for empowering patients, standardization of the medical decision process, overmedicalization, risks for individual freedom, and increasing social inequalities in health); *doctor protection* (fear of additional tasks and burden, doubt about the actionability of patient-gathered health data, risk for medical liability, dehumanization of the patient-doctor relationship, fear of increased drug prescription, and commodification of patient data).

Conclusions: A deep understanding of both the expectations and fears of GPs is essential to motivate them to recommend mHealth apps to their patients. The results of this study show the need to provide appropriate education and training to enhance GPs' digital skills. Certification of the apps by an independent authority should be encouraged to reassure physicians about ethical and data security issues. Our results highlight the need to overcome technical issues such as interoperability between data collection and medical records to limit the disruption of medical work because of data flow.

(JMIR Mhealth Uhealth 2021;9(3):e21795) doi:10.2196/21795

KEYWORDS

mobile applications; qualitative research; general practitioners; France; mobile phone

Introduction

Background

Mobile health (mHealth) technologies are apps or devices designed to deliver health information, collect clinical health data, provide real-time monitoring of patients' vital signs, and sometimes direct therapeutic interventions. The medical and health care literature presents mHealth apps as a means of empowering patients by providing them with medical and health-related information and education, improving their compliance with treatment, and helping them manage their own health [1]. Many patients consider that mHealth apps aimed toward enabling patients to facilitate the self-management of their illnesses [2], especially in the case of multimorbidity that complexifies patients' care needs, leading to difficulties in coping with illnesses and an increased burden of illness and treatment [3]. In addition, several clinical trials have shown a meaningful effect on health outcomes attributable to apps that mostly address diabetes, mental health, and obesity [4].

In France, as in most European countries, general practitioners (GPs) occupy a central place in the health care system. In 2014, more than 80% of French people aged 15 years and older had consulted a GP during the year [5]. Not only are GPs primary care providers and gatekeepers but they also manage and coordinate the health care pathways of their patients [6-8].

Multimorbidity, defined as the simultaneous coexistence of more than 1 chronic condition in a single individual, is increasing in primary care and leading to more complex medical care [9]. GPs play a central role in organizing care delivery for patients with multimorbidity [10] and managing complex decision making and providing patients with support to self-manage their illnesses [11]. To provide a comparison with another European country, in Scotland, about a quarter (23.2%) of the population has chronic multimorbidities [12]. The increase in multimorbidity does not only concern the older adults; by the age of 50 years, half of the Scottish patients had at least one morbidity in 2007 [12]. This leads us to believe that a large proportion of chronically ill patients should be concerned with effective tools or devices that could help both patients and their physicians to manage their illnesses, especially through mHealth technologies.

mHealth apps are available on smartphones or tablets. In 2019, almost 8 out of 10 French people had a smartphone, but only 44% of those aged 70 years and older had a smartphone [13]. In this population, the use of a smartphone to browse the internet decreases after 40 years of age: 85% (40-59 years), 69% (60-69 years), and 57% (70 years and older). In the end, only 26% of all people aged 70 years and older browse the internet on a mobile device; however, this rate has been increasing over the past year (+7 points for those aged 70 years and older and +10 points for those aged 60-69 years) [13].

In France, GPs can prescribe some connected devices, that is, these devices are reimbursed by the health insurance scheme (eg, glucometers such as Freestyle Libre). Currently, only 1 app (Moovcare) is reimbursed by the health insurance scheme and can be prescribed for the clinical follow-up of patients with lung cancer. Other stand-alone mHealth apps are not yet reimbursed; therefore, doctors can only advise them. However, will French GPs be ready to prescribe mHealth apps to their patients when these apps will be certified by health authorities?

Both a Swedish [14] and an Australian [15] study suggest that physicians tend to recommend health apps to their patients sparingly, even if they have a positive attitude and perceive an improvement in patients' self-management ability as the main benefit of health apps [14]. The lack of knowledge of effective apps seems to be the main barrier [15,16]. Dutch GPs seem to have a supporting or mixed attitude toward the integration of mHealth in primary care [17]. A German survey [18] conducted with 1070 GPs showed that "skeptical physicians"—who are concerned about data security, reliability of apps, legal questions, and additional burdens—are more numerous (44%) than the "open-minded physicians" (35%), who emphasized motivational and compliance advantages. However, a majority of the GPs interviewed perceived valuable application potential for health apps and positive contributions to health care and recovery. They are now in favor of recommending mHealth apps to patients, even if they had been reluctant to do so in the past.

Objectives

The aim of this study is to investigate French GPs' attitudes toward prescription or recommendation of mHealth apps or devices aimed toward patients by analyzing their perceptions and expectations of mHealth technologies. On the basis of a sociological approach and following previous research [19-21], this paper aims to provide an empirically grounded typology of the attitudes of French GPs toward the prescription of mHealth apps.

Methods

Design

This qualitative study, which was conducted from October 2018 to March 2019, involved semistructured face-to-face interviews and 2 focus groups (FGs) with GPs.

Ethical approval was obtained from the French Institute of Medical and Health Research Ethics Committee (IORG0003254 and FWA00005831) and the Institutional Review Board (IRB00003888; opinion number 18-499). All participants gave their consent after receiving both oral and written information about the study.

Study Participants and Recruitment

As French general practice is predominantly in the private sector [22], all participants were in private practice.

On the basis of a purposive sample strategy, 20 GPs working in South-East France were interviewed individually. To obtain various viewpoints about GPs' perceptions and expectations of mHealth apps aimed toward patients, we interviewed GPs who were at different stages in their career trajectory, worked in areas with different population densities, and worked in solo or group practice (single specialty or multispecialties; Table 1). As part of the recruitment of study participants, 90 physicians were initially contacted using a professional phone directory.

Of these, 12 agreed to participate in an interview, 6 refused to participate, and a large majority (72/90, 80%) did not give an answer (positive or negative). A total of 8 participants were recruited via snowball sampling. The recruitment process ended when the theoretical saturation point (defined later) was reached.

Table 1. General practitioners' characteristics.

Characteristics	Individual interviews (n=20)	Focus groups (n=16)	Total sample (N=36)
Age (years), n (%)			
<40	5 (25)	9 (56)	14 (39)
40-50	1 (5)	1 (6)	2 (6)
50-60	6 (30)	4 (25)	10 (28)
>60	8 (40)	2 (13)	10 (28)
Gender, n (%)			
Male	12 (60)	8 (50)	20 (56)
Female	8 (40)	8 (50)	16 (45)
Private practice context, n (%)			
Solo practice	5 (25)	1 (6)	6 (17)
Group practice	12 (60)	9 (56)	21 (28)
Multidisciplinary primary health care organization	3 (15)	6 (38)	9 (25)
Involved in medical training of residents, n (%)			
Yes	14 (70)	16 (100)	30 (83)
No	6 (30)	0 (0)	6 (17)

In addition, to debate about physicians' expectations for mHealth apps, 2 FGs were organized with 7 and 9 GPs each. All 16 participants were involved in university training for the GP residents. The first FG was conducted in Nice (South-East of France), whereas the second was held in Rouen (North-West of France; Table 1).

Procedure

Individual interviews were conducted by a medical sociologist (LA). She approached the participants by email and conducted face-to-face interviews at physicians' offices with willing participants for a duration of 28 to 87 minutes (median=40).

The FGs were organized in a university setting. Each FG was conducted by an experienced moderator (TB and MS), and the cumulative duration of the interviews was 132 minutes.

An interview guide was used to conduct the interviews and FG discussions. It covered the following areas: knowledge and interest in digital technologies, possible use of mHealth apps, experience of recommending mHealth apps or devices, effects on doctor-patient relationship, risks and benefits for the patient, doubts and fears, and effects on patient responsibility and autonomy. We used a FG approach designed to explore the stable opinions, norms, and group processes that arise within the group.

Interviews and group discussions were audio recorded and transcribed. All personally identifiable information was removed from the transcripts. Interviews and FG were identified with a code: E (for interview)+number, FG (for focus group)+city initial (R for Rouen, N for Nice)+age.

Data Analysis

A qualitative analysis of each interview and FG was conducted using grounded theory coding analysis [23]. Transcripts of the personal and collective interviews were hand coded using an iterative inductive process. The first step was initial coding, and segments of transcripts were coded into categories until the categories accounted for all the variations in the data. In the second step (focused coding), selective coding and relationships between the categories were refined and core categories were identified and arranged into broad emergent thematic categories. The third step (axial coding) aimed at identifying the core phenomenon, causal conditions, resultant strategies, context, and consequences [23].

To propose a researcher-constructed typology to simplify and "to reduce the complexity of the natural world by focusing attention on a usually small number of elements or issues of interest to the researcher" [24], we searched for an empirical dimension connected with our research object (GPs' perceptions of mHealth, expectations, concerns, and fears). We identified the dimension of the value given by GPs to the mHealth apps aimed toward patients. This empirical dimension has emerged as a central issue in our data and allowed to organize the data in new ways following 2 axes: (1) the value of the apps for patients (from high to low) and (2) the value of the apps for medical work (from high to low). Next, we reassembled the data into 4 themes (Table S1 given in Multimedia Appendix 1): (1) GPs' willingness: apps as additional resources to medical work, (2) GPs' skepticism or resistance, (3) high value of the apps for patient, and (4) less value of the apps and risks for patient.

Finally, we constructed a typology that is a theorizing process to understand and explain the phenomenon at hand, that is, the range of attitudes of GPs toward the possible prescription of mHealth apps. Considering the position of GPs' concerns and expectations on the value axes, 3 groups of attitudes or positions adopted by the GPs were identified and defined: *digital engagement*, *patient protection*, and *doctor protection*.

Interview transcripts were coded by LA and checked by AS. The FG transcripts were coded by TB and checked by DD and AS. To ensure the validity of the analysis, the following strategies were used: saturation and constant comparison. Data collection and analysis were conducted simultaneously, leading the purposeful sampling to confront the validity of analysis or to explore new concepts until no further new items emerged (theoretical concept saturation was reached). The data codes, categories, and themes were constantly checked, compared, and contrasted. Selective coding and relationships between the categories were discussed within our interdisciplinary research team that included a medical anthropologist (AS), a medical sociologist (LA), and 3 GPs (DD, TB, and MS) who are involved in research in general practice and training of junior GPs. These findings were also examined in relation to the existing literature to identify any inaccuracies and/or misinterpretations.

Results

Attributes and Types

Considering the value assigned to mHealth apps by participants and their willingness or resistance to prescribe, we identified 3 positions: *digital engagement*, *patient protection*, and *doctor protection*. However, these categories are ideal types of attitudes and not of people. In other words, each GP interviewed tended to adopt one type of attitude but had an ambivalent discourse regarding specific aspects of the prescription of mHealth apps or the patients' use of apps.

We detail these 3 positions by giving their characteristics, which we illustrate with a few short relevant verbatims. In addition, the reader may find other illustrative excerpts from interviews classified by themes during the final coding in Table S1 in Multimedia Appendix 1.

Digital Engagement

Some of the GPs adopted a rather favorable posture for mHealth and perceived mHealth apps as additional resources. Most of them were familiar with the tools; they used professional apps and had tested some mHealth apps:

> *One that is very often recommended for chronic lumbago is the application Activ'Dos. I know that one well because I have used it myself. It is rather well done. (...) I tested it on myself, so I advise it for my patients and show them how to use it.* [E19]

Although our sample is not a representative one, it is instructive to stress that among the 20 GPs interviewed individually, 10 used professional apps, 11 knew of one or more mHealth apps aimed toward patients, and 6 had previously advised an mHealth app to their patients. Most of them reported a lack of knowledge or familiarity with these tools and even a lack of interest. However, they expressed a desire to gain more digital skills.

The main perceived benefit was the facilitation of medical work and follow-up care. Participant GPs stressed the reliability and objectivity of measurements allowed by the app, its traceability, and the possibility of a history of measurements that facilitate medical work in terms of diagnosis and patient care plan:

> *[About blood pressure measurements] An application would be much better. I give them papers with charts to fill in, then they bring me their charts...which is just fine! I scan them and put them in the files, but if I had that on the Internet, it would be...much simpler.* [E15]

A patient's digital logbook is perceived to be more convenient than the usual paper logbook. From this point of view, digital tools are additional control devices that could facilitate better monitoring of patients, because this self-monitoring could be more intensive and occur in real time:

> *It should be a tool that facilitates the transmission of information from him to me. (...) It would be a bit like having the nurse visit the patient every day (...) with an eye on the patient's state of health on a daily basis. I think that would be useful.* [E10]

The case of diabetic patients was often cited as an example of the utility of the mHealth app, especially to help physicians in remote decision making. On the basis of the history of blood glucose tests transmitted via an app connected to the medical records, and without meeting the patient, the physician could adjust the therapeutic regimen and dosages:

> *The major interest is the communication of data (...) so the interest is double for the diabetic, (...) there is no need to go anywhere. All of the data are transferred to the doctor. He has, for example, a week of (blood sugar) curves and can remotely (...) adapt the insulin and the dosage and the therapeutic measures for the diabetic.* [E7]

From their point of view, the mHealth app could serve as an extension supporting previously practiced medical work for tasks such as relaying advice related to healthy food or physical activities, self-monitoring, and education. GPs interviewed presented mHealth apps aimed toward patients as complementary tools to medical practice; however, they stressed that an app should not replace the physician:

> *[mHealth apps] This would be to help the doctor on a daily basis, or to help the patient in the care and management of his illness, but certainly not to make medical diagnoses or to provide pseudodiagnoses (...) in the place of the doctor.* [E1]

They also perceived benefits related to patients' engagement in their care. From the point of view of physicians interviewed, mHealth apps could improve patients' compliance with treatment and provide patients with reliable health-related information and education. The availability of targeted, easy-to-access information could help them manage their own health and adopt a healthy lifestyle:

> *Sometimes when we give advice, we don't know what happens once they have gone home. If they have their application, it will support our advice about diet regimes, advice about care for certain chronic illnesses like diabetes... So, these are tools to help gain knowledge about their illness, to better understand the complications.* [E20]

Patient Protection

Although GPs perceived benefits in terms of patient involvement and engagement, they also put forth a protective posture vis-à-vis their patients and the possible risks such apps could engender. Some of the risks mentioned were related to smartphone use and were falling within the lay representations of the technology:

- Addiction to smartphones:

> *It depends on the tool and depends on the usage of that tool. There are uses which become addictive, which become stressful, and harmful.* [E11]

- Dangers of cell phone radiation:

> *What does all the radiations do? (...) they say that telephones can cause brain tumors? (...) We are not 100% sure of their safety, so the person who is going to go around all the time with applications that link data, isn't it potentially dangerous?* [E3]

- Distortion of ongoing relationships and cognitive changes:

> *The fact of having your telephone in hand all the time...it necessarily cuts off social interaction between people. And also, there are many cognitive changes when we are used to using apps from such a young age (...) we make less of an effort to remember things, less of an effort to search for information where it existed before, we have perhaps less critical distance from the sources we use (...) some sort of middle ground has to be found between using an application from time to time and without it severing all preexisting social ties.* [E19]

Another risk mentioned was related to the content of the apps that could provide irrelevant, unreliable, or nonevidence-based information:

> *There is always a risk that the information is not adapted to their case.* [E18]

From the point of view of the study participants, a certification performed by independent authorities, such as governmental agencies and health care professionals, could prevent the aforementioned risk:

> *I do not give advice to a patient if there is no scientific argument to back it up. I cannot suggest an app if it is not scientifically valid, or even independent, and if the protection of personal data is not guaranteed.* [E2]

They were also concerned about the security of the apps and had several fears about patients' data privacy and security:

> *An app, proposed by a Lab [pharmaceutical industry]: NO! I would not trust it! [the risk would be] targeting, collecting patient data in their favor, to promote their products (...) because I think that the Lab collects patient data because the patients enter the data. To what end will they use this data? I am also there to protect my patients' DATA, and I want to guide them in the use of applications that will not put their data protection at risk.* [FGN-30.2]

This protective posture was also fueled by the perception of another category of risks related to the management of patient care. From these GPs' point of view, self-medication or self-management could lead to patient isolation, increase patients' anxiety, or give patients a false sense of security:

> *The objective is not to collect data but to know how to analyze the data in a relevant and pertinent way. (...) it really must be standardized, otherwise there is a false sense of security: "I feel all alone."* [E8]

Moreover, health apps are not seen as relevant tools for empowering patients because, as a participant said, "those who are going to get involved are patients who are already involved" (E17), whereas those who had little self-involvement in the management of their health will not become more empowered by mHealth apps:

> *We aren't going to gain much. It will over-empower those who were already too much and...not change much.* [E8]

They also pointed to the risk of the standardization of the medical decision-making process based on patient monitoring with apps, which is opposed to a family practice. This standardization could lead to overmedicalization:

> *We can adapt these software programs all we like, I am not sure that we will ever be able to obtain the intuitive perspective that is the transcription of countless information that we have accumulated through experience and knowledge about our patients.* [E20]

They feared that apps might increase normative injunctions to achieve perfect health. In addition, they feared that patient autonomy requirements would put too much responsibility on patients and decrease their individual freedom:

> *Surveillance and self-management is all right but we shouldn't be obsessed with all these settings. (...) There is too much information, I find, from the apps, the television, advice... (...) We have the right to cheat a bit, but I have the impression that we must live in a world that is more and more perfect.* [E9]

Finally, some of them suggested that apps could intensify social inequalities in health, being prescribed only to certain patients, depending on their income level, their level of digital literacy, and/or their linguistic ability:

> *What counts also, because I work in a difficult neighborhood, is applications that are adaptable–so, according to the languages the people speak, their level of literacy, because I am afraid that some apps*

> *are too specialized, and that deepens social inequalities of health!* [FGN-30.1]

Doctor Protection

In response to the issues and challenges of the prescription of mHealth concerning medical practices, some GPs adopted an attitude that aims to protect their work and their professional position. They view the prescription of apps as an additional task. For some, giving related explanations would be too time consuming in the context of the current constrained duration of the medical encounter:

> *There are a lot of parameters to set up...For example, I tried the app for the pill: it really got on my nerves because by the time I found the right pill, I was running late! It isn't manageable during a consultation! It blocks everything! For me, it should be a PLUS, and not take me more than 5 minutes, or...10 minutes; otherwise, I won't be able to manage! (...) For an app to work, you have to enter the name, the age, and a certain number of other things that will take up too much of MY TIME! That is what I fear!* [FGN-60]

Apps would offer scattered information that might generate a great quantity questions from patients:

> *If the patient comes with more information (...) if he has read up, he gets off track. He has even more questions. (...) Raising questions is a very good thing because patients have to, of course, be informed...but at their level (...) during the consultation, there are often several reasons for the consultation, we have to think about performing cancer screening from the age of 50, we have to think about vaccinations... this is just adding yet another thing to think about (...) during the consultation, we just don't have the time.* [E10]

Likewise, patients' use of health apps might space out medical follow-up that would lessen the doctor's acquaintance with his or her patient and increase the time necessary to know the important events that occurred in the meantime:

> *So they are more independent and self-reliant which is good, but when they come back to see us it will be a waste of time because what have they done during all this time? Why did they do this or that? The patients are less closely monitored.... So of course this will forcibly be more difficult to manage.* [E18]

The management of patient-gathered health data (such as patient-reported outcomes or patient data automatically generated by apps) will create extra work and increase the workload, especially because of the noninteroperability between the apps and the electronic medical record:

> *I REFUSE to be INUNDATED with "patient" data that arrive every evening and that I am supposed to have seen, with people who will come in and say "Have you seen my curves? What do you think about that?" That, that WOULD BE AWFUL!!! It would be the last straw! (...) That (the data flow) shouldn't be an extra burden (...) So, I can imagine to what point the apps, if they provide a constant flow of information when the patient isn't even here, would really make things difficult for me.* [FGR-50]

GPs are dubious about the value of patient-gathered health data compared with other clinical data to improve patient care and care planning and about the actionability of such data in the clinical realm:

> *I am bothered when people come and show me the results of their Freestyle app (...) I am lost with the averages in the result curves which I could care less about, and then we have to regulate the insulin...NO! the ONLY THING I AM INTERESTED IN is the morning blood sugar level!! I couldn't care less about the rest!* [FGR-50]

They raised concerns of medical liability associated with remote monitoring:

> *The fear is that you prescribe mobile apps to all of your patients (including those with high blood pressure) and then, one night you are snug in bed and you receive a text message "Mr. So-and-So has a systolic of 200!!" [Laughter] ... you don't sleep a wink all night!* [FCN-40]

Moreover, they emphasized that health apps should not be a substitute for diagnostic tasks:

> *The patient mustn't go imagining that that (apps) will replace us. (...) it is just a digital tool and doesn't have a global vision, a vision of everything and anything that could interfere with this or that... It must be a complement to what already exists; it must not replace the doctor's point of view.* [E15]

mHealth apps are thought to be a disruptive technology that could alter care practices by reinforcing the purely technical aspects of medicine. Some GPs questioned the risk of transferring their role to the tools. Dehumanization of relationships is feared: mHealth apps could limit the specificities of general medical practice, such as negotiating with the patient's follow-up plan and its adjustments, and might reduce interaction with patients and the tracking of emotions that are important to medical diagnosis and follow-up:

> *We might even be tempted to do that. (...) To delegate the work (...) (but) the empathy that we put into the explanation (of the illness), that is capital for the patient (...) we will have to learn to use that, but without forgetting our role, that does not discharge us from this role of explaining and sharing, especially (...) because that contact, when we no longer have it will be the moment when we become technicians and we will only have to apply algorithms.* [E20]

Another concern was the fear of increased drug prescriptions and health care expenses, especially if health apps are developed by the pharmaceutical industry:

> *There is an app that is officially provided by a patient organization, but in reality a lab (pharmaceutical industry) is behind it and financing it (...) The app is*

> *for the simple detection of AMD [age-related macular degeneration] (...) it is financed by the lab that supplies the products. The injection costs 800 euros to prevent AMD!* [E4]

That is why they were favorable to certification by independent authorities, such as governmental agencies and the health care professional community:

> *I would prescribe them (the apps) more readily if there were some authority that certified them and had the reputation of being independent...* [E14]

Finally, they fear the exploitation of people's personal information, the commodification of patient data, and a new form of population monitoring:

> *The diffusion of the mass (of data) is something I fear much more. And that after that, we would treat certain sub-populations, in a statistical sense of the term, sub-populations with differentiated risk-management, that is a huge risk.* [E8]

Discussion

Principal Findings

On the basis of the perceptions and expectations of mHealth apps aimed toward patients of a sample of 36 French GPs, our study offers a typology that makes it possible to account for the different attitudes of GPs with regard to the recommendation of apps. Rather than discussing each category of attitudes, we analyzed the GPs' fears, concerns, and tensions highlighted by the typology that can explain the resistance, on some of their parts, to recommend or prescribe mHealth apps to patients.

The first source of tension emerges in the approach of the doctor-patient relationship and the medical work of care management. mHealth apps are frequently presented in medical and public health literature as a means of empowering patients [25], and several GPs are sensible to this argument to prescribe mHealth. Several qualitative studies suggest that patients perceive mHealth apps as a means of self-monitoring and self-management of their health and an aid in gaining control and autonomy over their own health and health care [2]. Although the GPs interviewed approved patient engagement in their health care, they feared risks that did not seem to have been disclosed by previous studies. They pinpointed risks related to doctor-patient relationships (dehumanization), patient follow-up (isolation, anxiety, self-medication, false sense of security, numerical standards for normalcy, and disappearance of doctor mediation in the interpretation of data), and medical practice (substitution for medical diagnosis, overmedicalization, standardization of follow-up and prescription, and reduction of professional autonomy). Finally, like their "skeptical" German colleagues [18], the GPs interviewed stressed security, privacy, and confidentiality concerns about patient-generated data.

mHealth can generate tensions if the tool provides patient information that the GP does not need for his or her expertise or if the flow of information disrupts medical work, reduces valuable medical time, or increases the workload. This is consistent with a previous study [18]. To manage care and follow-up, GPs require different information about their patients. Therefore, mHealth apps can support medical practice in generating patient data [1]. However, as shown by previous studies, doctors "feel more comfortable about data that they have collected themselves about the patient (...) rather than those collected by patients themselves on their own initiative" [26]. To consolidate their expertise (medical decision making and plan care), they expect data "valence of actionability" [27], which refers to patient-generated data integration in clinics and the physician's responsibility in subsequent medical interventions.

The GPs interviewed stressed the lack of guidelines for patients' use of mHealth apps and the need for independent certification to ensure quality and trustworthiness of the apps before recommending them, as suggested by previous studies [14,15]. Certification of mHealth apps could be a way to increase GPs' confidence in mHealth apps and promote the digital engagement of physicians. However, the huge production of new mHealth apps renders certification tasks more complex [28].

Our typology underlines tensions in GP's medical work that would be induced by patients' use of mHealth apps. However, these tensions could be experienced by GPs differently, depending on national health care contexts and considering the various patient requirements and expectations that may differ according to the diverse organizations of health care systems in other countries. For instance, some mHealth apps may be alternatives to medical consultation in developing countries where access to health care is restricted or in industrialized countries where care costs are substantial and for patients living far from health care services. On the other hand, a solidarity-based public health insurance system, such as the French system that covers 70% to 100% of primary care costs, does not favor patient self-care. As a result, prescribing mHealth apps is more likely to be perceived by GPs as a supplementary workload. However, the globalized mHealth technologies market proposes the use of mHealth apps without distinguishing between the different expectations of each relevant actor in each country.

Finally, transversal to these different attitudes, in our study, most GPs consider the smartphone to be an integral part of daily lives of many people. Not all physicians interviewed were entirely convinced that mHealth apps have the potential to revolutionize the management of disease. However, they regarded the introduction of mobile apps as an inevitable development in the field of medicine and one that could be leveraged to change patients' behavior.

As we have emphasized, each GP interviewed could have an ambivalent discourse with regard to specific aspects of the prescription of mHealth apps or patients' use of apps. Some sociological approaches propose a nuanced understanding of users' attitudes toward digital health [29]. Rather than the classical opposition found in the literature between "resistance to" [30-33] and "acceptance of" [34-37] digital health technologies, Marent et al [29] prefer to use the concept of ambivalence as a "means of facilitating and capturing contradictions and tensions in accounts of digital health" [29]. According to these authors, "ambivalence is the simultaneous experience of two (ambi) opposing orientations or values

(valences)" that "indicates an oscillation or tension between opposite poles of feeling and thinking," at an individual or collective level. In the case of the physicians interviewed, we hypothesize that this ambivalence is generated by a tension between the desire to see patients' commitment to care progress and the fear of seeing their medical authority challenged [26]. To alleviate this tension, some GPs interviewed adopted a paternalistic stance [26] by highlighting their medical responsibility to protect patients and medical ethics [38].

From a daily work perspective, our findings could help GPs to identify the category of attitudes with regard to the recommendation of apps in which they fall and to make them aware of their ambivalence. They could then anticipate strategies to lessen their personal constraints on practices such as task delegation, personal training, adequate medical record software choice, production of information devices for patients, and identification of apps that match their values and expectations. Subsequently, they could slowly but steadily start prescribing mHealth apps to targeted patients at targeted times.

Limitations

The main limitation of this study is the high rate of nonrespondents (72/90, 80%). Moreover, the recruitment process led to a high number (30/36) of GPs involved in medical training. This overrepresentation may be a bias because respondents, especially physician trainers, could be aware of mHealth and more willing to participate in the study. It is possible that the nonrespondents were not interested in mHealth and, therefore, were even less concerned about their prescription.

However, the "thick description" [39] of the various GPs' postures and their latent meanings revealed by sociological analysis make our typology transferable to other national contexts [39].

Recommendations for Implementation

As GPs are leading providers of health information and advice to patients, a deep understanding of both expectations and fears of GPs is essential to motivate them to recommend mHealth apps to their patients. To this end, there is a need to provide appropriate education and training to enhance the digital skills of GPs [15,40]. Our findings show that certification of the apps by an independent authority should be encouraged to reassure physicians about ethical and data security issues [14,15,28]. Our findings also highlight the need to overcome technical issues such as the interoperability between data collection and medical records to limit the disruption of medical work by the data flow. It could be useful to reorganize the medical work, with consultation devoted both to the prescription of apps and the interpretation of the patient-generated data, to lift economic barriers and to preserve personalized management of patient care.

We can anticipate that the availability and assistance of self-monitoring apps in the context of the response to the COVID-19 epidemic to follow infected patients [41,42] could impact the willingness of GPs to prescribe mHealth apps to their patients.

Conclusions

The results obtained using a sociological analysis clearly reveal the tight imbrication of social and cultural factors underlying the attitudes of GPs toward prescriptions of mHealth apps. Our analysis could be informative for medical teachers, public health decision makers, and all professionals grappling with questions of integrating mHealth in medical practices.

Acknowledgments

This work was supported by the French Agence Nationale de la Recherche (ANR) under grant ANR-17-CE19-0027 (Projet ApiAppS). The authors would like to thank all GPs who agreed to participate in this study. The authors would also like to thank Jessica Hackett for revising the English manuscript and the reviewers for their relevant and helpful comments.

Authors' Contributions

All authors contributed to the conception and design of the study. Data collection was performed by LA, TB, and MS. All authors contributed to data analysis. AS-E wrote the first draft of the manuscript, and all authors commented on the previous versions of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Illustrative quotes.
[DOCX File , 25 KB - mhealth_v9i3e21795_app1.docx]

References

1. Topol E. The Patient Will See You Now. New-York, USA: Basic Books; 2016:384.
2. Vo V, Auroy L, Sarradon-Eck A. Patients' Perceptions of mHealth Apps: Meta-Ethnographic Review of Qualitative Studies. JMIR Mhealth Uhealth 2019 Jul 10;7(7):e13817 [FREE Full text] [doi: 10.2196/13817] [Medline: 31293246]

3. Runz-Jørgensen SM, Schiøtz ML, Christensen U. Perceived value of eHealth among people living with multimorbidity: a qualitative study. J Comorb 2017;7(1):96-111 [FREE Full text] [doi: 10.15256/joc.2017.7.98] [Medline: 29359124]
4. Byambasuren O, Sanders S, Beller E, Glasziou P. Prescribable mHealth apps identified from an overview of systematic reviews. NPJ Digit Med 2018;1:12 [FREE Full text] [doi: 10.1038/s41746-018-0021-9] [Medline: 31304297]
5. Célant N, Guillaume S, Rochereau T. Enquête santé européenne: enquête santé et protection sociale (EHIS-ESPS) 2014.: Les Rapports de l'IRDES; 2017 Sep. URL: https://tinyurl.com/2x74x4wp [accessed 2021-02-16] [WebCite Cache ID https://tinyurl.com/2x74x4wp]
6. Macinko J, Starfield B, Shi L. The contribution of primary care systems to health outcomes within Organization for Economic Cooperation and Development (OECD) countries, 1970-1998. Health Serv Res 2003 Jun;38(3):831-865 [FREE Full text] [Medline: 12822915]
7. Starfield B. Primary care and health. A cross-national comparison. JAMA 1991;266(16):2268-2271. [Medline: 1920727]
8. Starfield B, Shi L, Macinko J. Contribution of primary care to health systems and health. Milbank Q 2005;83(3):457-502 [FREE Full text] [doi: 10.1111/j.1468-0009.2005.00409.x] [Medline: 16202000]
9. Uijen AA, van de Lisdonk EH. Multimorbidity in primary care: prevalence and trend over the last 20 years. Eur J Gen Pract 2008;14 Suppl 1:28-32. [doi: 10.1080/13814780802436093] [Medline: 18949641]
10. Wallace E, Salisbury C, Guthrie B, Lewis C, Fahey T, Smith SM. Managing patients with multimorbidity in primary care. BMJ 2015 Jan 20;350:h176. [doi: 10.1136/bmj.h176] [Medline: 25646760]
11. Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness. JAMA 2002 Oct 9;288(14):1775-1779. [Medline: 12365965]
12. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet 2012 Jul 7;380(9836):37-43 [FREE Full text] [doi: 10.1016/S0140-6736(12)60240-2] [Medline: 22579043]
13. Enquête sur la diffusion des technologies de l'information et de la communication dans la société française en 2019. BAROMÈTRE DU NUMÉRIQUE 2019. 2019. URL: https://www.credoc.fr/publications/barometre-du-numerique-2019 [accessed 2020-04-17] [WebCite Cache ID https://www.credoc.fr/publications/barometre-du-numerique-2019]
14. Zhang Y, Koch S. Mobile health apps in Sweden: what do physicians recommend? Stud Health Technol Inform 2015;210:793-797. [Medline: 25991263]
15. 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]
16. Sezgin E, Özkan-Yildirim S, Yildirim S. Investigation of physicians' awareness and use of mHealth apps: A mixed method study. Health Policy and Technology 2017 Sep;6(3):251-267. [doi: 10.1016/j.hlpt.2017.07.007]
17. Bally ELS, Cesuroglu T. Toward Integration of mHealth in Primary Care in the Netherlands: A Qualitative Analysis of Stakeholder Perspectives. Front. Public Health 2020 Jan 15;7. [doi: 10.3389/fpubh.2019.00407]
18. Wangler J, Jansky M. Welchen Nutzen bringen Gesundheits-Apps für die Primärversorgung? Ergebnisse einer Befragung von Allgemeinmedizinern. Präv Gesundheitsf 2020 Jun 26. [doi: 10.1007/s11553-020-00797-7]
19. Hammer RP, Burton-Jeangros C. Tensions around risks in pregnancy: A typology of women's experiences of surveillance medicine. Social Science & Medicine 2013 Sep;93:55-63. [doi: 10.1016/j.socscimed.2013.05.033]
20. Lebeau J, Cadwallader J, Vaillant-Roussel H, Pouchain D, Yaouanc V, Aubin-Auger I, et al. General practitioners' justifications for therapeutic inertia in cardiovascular prevention: an empirically grounded typology. BMJ Open 2016 May 13;6(5):e010639. [doi: 10.1136/bmjopen-2015-010639]
21. Pfaller L, Hansen SL, Adloff F, Schicktanz S. 'Saying no to organ donation': an empirical typology of reluctance and rejection. Sociol Health Illn 2018 Jun 28;40(8):1327-1346. [doi: 10.1111/1467-9566.12775]
22. Bouet P. Atlas de la démographie médicale en France. Situation au 1er janvier 2020. Paris: CONSEIL NATIONAL DE L'ORDRE DES MEDECINS; 2020. URL: https://www.conseil-national.medecin.fr/sites/default/files/external-package/analyse_etude/1grhel2/cnom_atlas_demographie_medicale_2020_tome1.pdf [accessed 2021-02-16] [WebCite Cache ID https://tinyurl.com/8ykhs9ws]
23. Charmaz K. Constructing Grounded Theory. London: SAGE Publications Ltd; 2014:416.
24. Schneider JW, Conrad P. Medical and sociological typologies: The case of epilepsy. Social Science & Medicine 1981 May;15(3):211-219. [doi: 10.1016/0271-7123(81)90004-3]
25. Lupton D. M-health and health promotion: The digital cyborg and surveillance society. Soc Theory Health 2012 Jun 27;10(3):229-244. [doi: 10.1057/sth.2012.6]
26. Lupton D. Digital Health: Critical and Cross-Disciplinary Perspectives. London, New-York: Routledge; 2018.
27. Fiore-Gartland B, Neff G. Communication, mediation, and the expectations of data: data valences across health and wellness communities. Int J Commun 2015;9:1466-1484 [FREE Full text]
28. CARE-INSIGHT, OPUSILINE. E-santé : faire émerger l'offre française en répondant aux besoins présents et futurs des acteurs de santé. Paris: PIPAME; 2016. URL: https://tinyurl.com/3c7jhd4n [accessed 2021-02-16] [WebCite Cache ID https://www.entreprises.gouv.fr/files/files/directions_services/etudes-et-statistiques/prospective/Numerique/2016-02-Pipame-e-sante.pdf]

29. Marent B, Henwood F, Darking M. Ambivalence in digital health: Co-designing an mHealth platform for HIV care. Social Science & Medicine 2018 Oct;215:133-141. [doi: 10.1016/j.socscimed.2018.09.003]
30. Timmons S. Nurses resisting information technology. Nurs Inq 2003 Dec;10(4):257-269. [doi: 10.1046/j.1440-1800.2003.00177.x]
31. May C, Gask L, Atkinson T, Ellis N, Mair F, Esmail A. Resisting and promoting new technologies in clinical practice: the case of telepsychiatry. Social Science & Medicine 2001 Jun;52(12):1889-1901. [doi: 10.1016/s0277-9536(00)00305-1]
32. Devan H, Godfrey HK, Perry MA, Hempel D, Saipe B, Hale L, et al. Current practices of health care providers in recommending online resources for chronic pain self-management. J Pain Res 2019;12:2457-2472 [FREE Full text] [doi: 10.2147/JPR.S206539] [Medline: 31496788]
33. Hennemann S, Beutel ME, Zwerenz R. Ready for eHealth? Health Professionals' Acceptance and Adoption of eHealth Interventions in Inpatient Routine Care. J Health Commun 2017 Mar;22(3):274-284. [doi: 10.1080/10810730.2017.1284286] [Medline: 28248626]
34. Davis FD. User acceptance of information technology: system characteristics, user perceptions and behavioral impacts. International Journal of Man-Machine Studies 1993 Mar;38(3):475-487. [doi: 10.1006/imms.1993.1022]
35. Puri A, Kim B, Nguyen O, Stolee P, Tung J, Lee J. User Acceptance of Wrist-Worn Activity Trackers Among Community-Dwelling Older Adults: Mixed Method Study. JMIR Mhealth Uhealth 2017 Nov 15;5(11):e173 [FREE Full text] [doi: 10.2196/mhealth.8211] [Medline: 29141837]
36. Dixon DR, Dixon BJ. Adoption of information technology enabled innovations by primary care physicians: model and questionnaire development. Proc Annu Symp Comput Appl Med Care 1994:631-635 [FREE Full text] [Medline: 7950004]
37. Venkatesh V, Davis F. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Management Science 2000;46(2) [FREE Full text] [doi: 10.1287/mnsc.46.2.186.11926]
38. Nye R. Médecins, éthique médicale et État en France 1789-1947. Le mouvement social 2006;214:19-26. [doi: 10.3917/lms.214.0019]
39. Guba E, Lincoln Y. Epistemological and Methodological Bases of Naturalistic Inquiry. Educational Communications and Technology Journal 1982;4(30):363-381. [doi: 10.1007/0-306-47559-6_19]
40. Ward R, Stevens C, Brentnall P, Briddon J. The attitudes of health care staff to information technology: a comprehensive review of the research literature. Health Info Libr J 2008 Jun;25(2):81-97 [FREE Full text] [doi: 10.1111/j.1471-1842.2008.00777.x] [Medline: 18494643]
41. Perlman A, Vodonos Zilberg A, Bak P, Dreyfuss M, Leventer-Roberts M, Vurembrand Y, et al. Characteristics and Symptoms of App Users Seeking COVID-19–Related Digital Health Information and Remote Services: Retrospective Cohort Study. J Med Internet Res 2020 Oct 20;22(10):e23197. [doi: 10.2196/23197]
42. Mehl A, Bergey F, Cawley C, Gilsdorf A. Syndromic Surveillance Insights from a Symptom Assessment App Before and During COVID-19 Measures in Germany and the United Kingdom: Results From Repeated Cross-Sectional Analyses. JMIR mHealth and uHealth 2020;8(10):e21364. [doi: 10.2196/preprints.21364]

Abbreviations

FG: focus group
GP: general practitioner
mHealth: mobile health

Edited by L Buis; submitted 25.06.20; peer-reviewed by P van der Keylen, M Holter; comments to author 22.09.20; revised version received 10.11.20; accepted 08.01.21; published 04.03.21.

Please cite as:
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
URL: https://mhealth.jmir.org/2021/3/e21795
doi:10.2196/21795
PMID:33661123

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org