JMIR mHealth and uHealth

Impact Factor (2024): 5.4
Volume 10 (2022), Issue 6 ISSN 2291-5222 Editor in Chief: Lorraine Buis, PhD, MSI

Contents

Viewpoints

Reviews

Original Papers

Corrigenda and Addendas

Viewpoint

Viewing Mobile Health Technology Design Through the Lens of Amplification Theory

Beza Merid[1], PhD; Maria Cielito Robles[2], BSc; Brahmajee K Nallamothu[3], MD, MPH; Mark W Newman[4], PhD; Lesli E Skolarus[2], MD

[1]School for the Future of Innovation in Society, Arizona State University, Tempe, AZ, United States
[2]Department of Neurology, University of Michigan Medical School, Ann Arbor, MI, United States
[3]Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI, United States
[4]School of Information, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Beza Merid, PhD
School for the Future of Innovation in Society
Arizona State University
1120 South Cady Mall
Tempe, AZ, 85281
United States
Phone: 1 (480) 727 8787
Email: Beza.Merid@asu.edu

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2022/6/e40273

Abstract

Digital health interventions designed to promote health equity can be valuable tools in the delivery of health care to hardly served patient populations. But if the design of these technologies and the interventions in which they are deployed do not address the myriad structural barriers to care that minoritized patients, patients in rural areas, and patients who have trouble paying for care often face, their impact may be limited. Drawing on our mobile health (mHealth) research in the arena of cardiovascular care and blood pressure management, this viewpoint argues that health care providers and researchers should tend to structural barriers to care as a part of their digital health intervention design. Our 3-step predesign framework, informed by the Amplification Theory of Technology, offers a model that interventionists can follow to address these concerns.

(JMIR Mhealth Uhealth 2022;10(6):e31069) doi:10.2196/31069

KEYWORDS

mHealth; digital health; cardiovascular disease; high blood pressure; structural barriers to health; racial health disparities; Amplification Theory of Technology

Introduction

Heart disease is a leading cause of death in the United States, killing roughly 655,000 Americans each year [1]. It also represents a disproportionate harm to minoritized people, who often face structural barriers to health including poor access to emergency medical services and treatment, insurance coverage, healthy foods, and safe environments for physical activity [2,3]. Efforts to monitor and prevent heart disease focus on the prevalence of key risk factors—including uncontrolled blood pressure and low-density lipoprotein cholesterol, a history of smoking, physical inactivity, and poor diet—and the role that health care providers and patients themselves can play in eliminating or minimizing their effect on patient and population health [4]. These risk factors in particular are prominent opportunities for intervention because, unlike other sources of risk such as age and family history, they are considered "modifiable" and, thus, an opportunity for providers to prevent disease and for patients to take action to secure their own health.

The relationship between providers and patients here revolves around the implementation of disease prevention strategies that are both effective in reducing morbidity and mortality while also being achievable within the resource constraints that shape

health care delivery and patients' daily lives. These strategies, in other words, try to offer practical solutions to address health care needs using tools, technologies, and means of communication that should be widely available to the populations they seek to serve. For example, researchers and providers studying racial health disparities in cardiovascular disease treatment and outcomes use SMS text messaging to facilitate communication between providers and patients; electronic home blood pressure monitors to enable the tracking of trends in blood pressure readings over time; and wearable devices such as the Fitbit and Apple Watch to monitor health metrics such as heart rate, exercise, and cardiac electrical activity [5,6]. These are used because they rely on technologies that are both accessible to patients on the consumer technology market and, with regard to the use of SMS text messages and activity trackers on smartphones, they make use of functions that are native to these devices and easy for users to incorporate in their daily lives.

While this approach to using digital health technologies to address modifiable risk factors for disease is an important modality of care, this viewpoint argues that *access* to these technologies does not guarantee the ability to afford or sustainably use them; it is merely one precondition of technology use that providers and researchers should consider when designing technological interventions to address patient needs. Equitably designed digital health interventions must also account for structural determinants of health that may shape how patients of different races, ages, and socioeconomic status, among other characteristics, would fare when encountering these interventions. This paper provides a predesign framework that interventionist health services researchers can implement, *prior to* deploying their digital health interventions, to think about facilitators of and barriers to technology use among patients whose resource constraints may shape their capability, opportunity, or motivation to address modifiable cardiovascular disease risk factors. We conclude by providing a case study where we apply this model in our ongoing work in this space.

Techno-Optimist Versus Techno-Pessimist Views of Mobile Health Interventions

Despite what the ubiquity of technologies such as smartphones and wearable devices might suggest about the promise and value of new technologies in our ongoing efforts to limit the disproportionate harm of cardiovascular disease on minoritized communities, their widespread commercial availability belies a fundamental tension about what we believe technology can do to address such disparities. This tension, broadly speaking, is between what we can call techno-optimist and techno-pessimist views of technology [7]. The former tends to view technology itself as additive or transformative, presuming that access to a given device is enough to create a desired change within the lives of its imagined users; the latter tends to believe that, without the provision of supportive infrastructures and attention to users' specific needs and barriers to use, technology itself may simply amplify existing inequities in access or opportunity. This acknowledgement of the need to think reflexively about technology and what we believe it can do is a central tenet of the "Amplification Theory of Technology," which calls on interventionist researchers to account for how the social conditions in which technologies are deployed fundamentally shape how—and if—they can be used.

In the context of health broadly, and mobile health (mHealth) interventions in particular, this theory explains how technologies can amplify adverse social determinants of health if they are not designed and deployed in a manner that is congruent with users' capabilities to use them. Many researchers and providers tend to align themselves toward the techno-pessimist position, worrying, for example, about a widening digital divide and the risk of creating "intervention-generated inequalities" [8]. This is an important concern, and it should be used to inform digital health technology and intervention design. The model for mHealth research that we propose here adopts this theory, and it asserts that techno-pessimism and a continued effort to develop technologies that address patients' needs are not only compatible positions for us to hold, but also part of a requisite relationship with technology itself.

Amplification Theory of Technology

We argue that this Amplification Theory of Technology, as formulated by Kentaro Toyama [7], should inform our efforts in designing mHealth interventions and, critically, the work we can do to ensure the safe and sustained use of these technologies by patients in underserved communities. The Amplification Theory described in Toyama's work makes 3 assertions.

First, it argues that technology cannot function as a substitute for institutional capacities or human intent that is missing among stakeholders or environments where an intervention is to be deployed; this is because technology is not a fixed force that, on its own, causes certain kinds of social change [7]. Such interventions require a scaffolding of social, political, and technological infrastructures to support the equitable deployment of a given technology.

Second, the theory argues that technology tends to amplify existing inequalities. Simply making a technology *accessible* to underserved populations will not ensure that the technology is usable among these populations, and will certainly not address structural conditions such as political and social marginalization, or a differential distribution of lifesaving resources. In contrast to a theory of technology that presupposes either a positive or a negative directionality of effect as a fixed impact, the Amplification Theory of Technology argues that technology is merely a tool "that multiplies human capacity in the direction of human intent" [7].

Third, the theory argues that technologies are most effective when they amplify successful intervention efforts with existing institutional capacity and intent to foster positive change, rather than attempting to fix or substitute for "missing institutional elements" [7]. Technologies can have both positive and negative effects, because they are magnifiers of human intent and capacity. This framing is in contrast to a view that might posit that universal access to a technology would function as a silver bullet for social problems.

We are interested in this kind of direct investment in human capacity and opportunity to use technology as part of our community-engaged research in Flint and Ann Arbor, Michigan, where we are working with community members with hypertension to develop an mHealth intervention that promotes physical activity and nutrition to help control blood pressure and prevent heart disease and stroke. We illustrate here how the development of mHealth interventions through the lens of Amplification Theory can help to negotiate the tension between techno-optimist and techno-pessimist positions and, further, provide a road map that health care providers and researchers can use to design interventions that use technology as support and an amplification of a broader social intervention to address persistent health disparities.

mHealth Technologies: Promises and Limitations

For health care providers and researchers who address disease disparities across the diagnostic spectrum, mHealth interventions may offer a sense of great promise in their capacity to deliver and improve health care. The use of SMS text messaging and smartphone apps to educate patients, "nudge" behavior change [9], enable continuous health monitoring [10], provide access to patient health information, and facilitate patient-provider communication [11] can generate impactful new ways to support patients and promote health. Building on existing efforts by patients to involve themselves in the management of their care and in decision-making processes, these interventions can enable patients to become "digitally engaged" [12] by adopting new media technologies that facilitate self-management. These types of interventions are informed by surveys and scholarship indicating that hardly reached populations—including minoritized people, people in rural areas, and people who may otherwise have trouble paying for health care—typically already have a mobile device such as a smartphone that can be used to this end [5,11]. The possibility of reaching these patients who may already have the capability and opportunity to use these technologies is exciting.

Attention to high rates of utilization of mobile devices among these populations is often a central focus of studies advocating for and evaluating the use of mHealth interventions, particularly among minoritized people [13]. Some researchers argue that, in the midst of a growing digital divide that exacerbates the harms of racism in health care, low-wage employment, and poor access to hospital facilities and providers, use and ownership of mobile devices can create a new means to self-manage disease risk and illness. They also suggest that mHealth interventions can offer a sense of social support to users by underscoring the value of health-promoting behaviors [14], thereby offering patients agency and a social infrastructure through which they can manage their health risks and outcomes.

However, these high rates of smartphone utilization do not tell the whole story. For example, in addition to documenting the near ubiquity of smartphone ownership, Pew Research reports that a plurality of smartphone owners say they use their smartphones, rather than a computer, to go online. But these data also indicate that there are notable demographic differences in this usage, including important distinctions by age. Pew notes that 60% of smartphone users between the ages of 18 and 29 and 51% of smartphone users aged 30-49 prefer using mobile devices to go online, as compared with 34% of users aged 50-64 and 28% of users aged 65 and older; conversely, 42% of smartphone users aged 50-64 and 44% of smartphone users aged 65 and older prefer using desktops, laptops, or tablets to access the internet, as compared with 22% of users aged 18-29 and 21% of users aged 30-49 [13]. These findings illustrate how the big picture of smartphone usage changes when we look at it with some granularity, in this case by comparing population segments by age. They also illustrate how, for example, an mHealth intervention that seeks to prioritize older adults would need to think carefully about the digital health strategies being employed, as well as the preferences and capabilities of the patients they hope to help, lest they exacerbate existing inequities in capability, opportunity, or motivation to use these devices [15].

We are taking a similar context-sensitive approach in our work with patients with hypertension in Flint and Ann Arbor to look beyond *access* to technology to consider the social, political, and economic conditions that may facilitate or prevent the use of our mHealth intervention. One problem with focusing this kind of work on access to technology is that it can situate the underutilization of digital technologies among particular populations as a problem for the patient, rather than as a problem for a health care system that disadvantages myriad patients within particular populations or social demographics. As Veinot et al [8] argue, this focus on individuals and individual-level health behaviors can be useful in triaging patients' emergent needs, but it misses an opportunity to work toward broader, structural solutions.

To move beyond framing this work around individual-level behavior change in digital health technology use, we deploy the COM-B (Capability, Opportunity, Motivation, Behavior) theory of behavior change to explain how structural conditions may impede patients' use of digital health technologies to prevent cardiovascular disease [16]. As Michie et al [16] explain, the COM-B model offers a theory of behavior change that accounts for social and community networks as well as general socioeconomic, cultural, and environmental conditions that can shape a person's capability, opportunity, or motivation to change a behavior. The focus of this model is on understanding behaviors—such as nonuse of a digital health technology recommended by a health care provider, for example—in its proper context, where the behavior can be situated as part of a broader social system. The 3 conditions necessary for behavior change, as explained by this model, are capability, opportunity, and motivation; in order to design digital health interventions that are likely to be successful, we argue, providers and researchers must think reflexively about how these conditions of behavior change may shape patients' relationships with the technologies we deploy. We discuss our application of this theory within our ongoing work on the Wearables in Reducing risk and Enhancing Daily Lifestyle (WIRED-L) study in the section that follows.

WIRED-L Study: Case Study

In our work at the WIRED-L Center, we engage in community-based participatory research to design an mHealth intervention that will assist patients with hypertension in lowering their blood pressure through increased physical activity and a healthy diet. Our approach to this work is informed both by this literature and our efforts to work with patient communities to understand what they would value in an mHealth intervention. Our framework for this intervention also includes a 3-step process, taking place prior to the deployment of our smartphone app, during which we apply the Amplification Theory of Technology to identify structural barriers to the use of our technological intervention as well as possible actions our collective research team can take to address these barriers and help facilitate the sustainable use of our mHealth intervention; we summarize these early stages of our community-engaged work below, and, following those details, share our 3-step process for thinking reflexively about technology in society in Table 1.

A primary goal of ours is to ensure that we are designing an intervention with—and not for—our community partners. As such, we are working with community leaders in community-based organizations focused on the health of older adults, community members who have participated in prior health studies in University of Michigan hospitals, and community members affiliated with the Community Ethics Review Board (CERB) [17] in Flint, Michigan, to discuss our shared vision for this work. The CERB is particularly important as it includes a group of community volunteers and leaders who conduct a review process to ensure that proposed research meets community needs, and that projects are sensitive to community culture.

Table 1. The 3-step process applying the Amplification Theory in addressing structural barriers to health technology use.

Steps	Sample questions	Examples of action
Step 1: Acknowledge the possibility of technology amplifying existing inequalities rather than transforming and immediately improving patient health	• Presuming access to a given technology, what do we know about users' capability or opportunity to use the technology at the center of our intervention? • Does any institutional capability to support this intervention already exist?	• Create a matrix documenting differential access or capability that may limit community partners' use of technological intervention. • Determine whether or not intervention relies on "myth of scale."
Step 2: Name structural, environmental, and social barriers that may prevent use within specific communities and among specific users	• Is the mHealth[a] intervention we are deploying accessible, affordable, and safe to use within our partner community? • What specific conditions may limit accessibility, affordability, and safety for users in this community? • What are the health effects of policy decisions such as "digital redlining," where internet service providers systematically exclude low-income neighborhoods from broadband access?	• Ask participants to identify environmental barriers to safe use of mHealth interventions (eg, lack of sidewalks and public park space as a barrier to physical activity interventions). • Identify existing limitations to local broadband internet connectivity, and articulate how structural barriers to information access can affect health.
Step 3: Identify and pursue coalitions to enact social, economic, and policy infrastructures needed to sustainably deploy interventions as designed	• Which providers, researchers, organizations, experts, and policymakers can help answer these questions? • How are we ensuring that community partners are active in this process, driving our inquiries and discussions about possible solutions? • What kind of funding is necessary to sustain the benefits derived from this intervention, and what can we do to secure it?	• Contact state legislature to call for allocation of public funding of broadband internet access for low-income patients and families who may benefit from mHealth intervention.

[a]mHealth: mobile health.

We interviewed community members with hypertension in both Flint and Ann Arbor to understand their capabilities, opportunities, and motivations to engage in cardiovascular disease risk factor reductions, to assess their use of technology in their daily lives, and their interest in a technological intervention to promote cardiovascular health. We also engaged in preliminary design workshops involving members of our team of health researchers and providers as well as our community partners. Our predesign research also included presentation storyboards shared with community members that created low-fidelity renderings of possible features that could be built into the mHealth app to assess and design toward our community partners' needs.

As we engage this mHealth design process that centers the needs of our patients, we have gained several insights. For instance, our Flint participants—who are predominantly Black and who reside in a majority Black city—report a lack of safe and accessible outdoor environments that facilitate physical activity for older adults; this is a finding borne out in research on interventions that seek to deploy technologies within hardly served populations, so while it does not represent a novel discovery in this context, we include this reflection here to note

that this barrier to physical activity is not experienced by our Ann Arbor participants—who are predominantly White and who reside in a majority White city—and to highlight the importance of using such disparities to inform an analysis of the political economy of health *as a part of* mHealth research. This acknowledgement of constraints in users' capability and opportunity to use a given technology, we argue, should directly inform the design choices we make before we deploy an intervention, and the community work we engage in after deployment to sustain the use of an intervention.

To that end, we present a 3-step process (Table 1), to be carried out prior to the deployment of a technological intervention, that providers and researchers can follow to ensure that the technologies they are designing do not inadvertently exacerbate existing inequalities, and to generate ideas about how they can also address the structural conditions that sustain these inequalities in the first place.

The first step in our predesign framework calls on providers and researchers to acknowledge the possibility that the technological interventions we design may amplify existing inequalities in rather that than transform or immediately improve patient health. We are drawing here on Toyama's [7] work on the Amplification Theory of Technology. Toyama [7] warns that we must look for the ways in which technology amplifies underlying human forces and social conditions. Asking questions about the assumptions we are making about technology, the directionality of influence we presume our technologies will have, and the differentials in access and motivation in user populations can enable us to address more directly these issues in our design, refinement, and deployment processes. The deliverable produced here should systematically document these beliefs and assumptions, and provide a baseline for reflection moving forward.

The second step in our predesign framework calls on providers and researchers to name the structural, environmental, and social barriers that may prevent the use of an mHealth intervention within a specific community and among specific users. This step is especially important in our contemporary moment when, following national and international attention to the disproportionate harms that police violence and poor access to quality health care have on the lives of Black and other minoritized people, providers and researchers are working to attune themselves to the health effects of structural racism; of course, this focus should always be a part of this research. In this step, we begin by considering questions of access to technologies, the affordability and sustainability of given devices, and whether or not they can be used safely and sustainably in a particular environment, and then we move onto situating these barriers within a structural context. We ask questions about the social forces that may shape individual behaviors, through both community engagement and feedback from the research team. The deliverable produced here should create a list of structural barriers to health that, in addition to reflections about the assumptions we are making about our technologies and patients' capability to use them, can help shape the decisions we make next.

The third step in our predesign framework calls on providers and researchers to identify and pursue coalitions of stakeholders who can help enact the social, economic, and policy infrastructures necessary to sustainably deploy these mHealth interventions as designed. If our work in Step 2, for example, identifies how "digital redlining," or the policies and investment decisions that "create and maintain class boundaries through strictures that discriminate against specific groups" [18], can impede patients' use of an mHealth intervention, what kinds of research, policy expertise, and investment decisions might we need to address these issues [19]? Likewise, as we are working to identify relevant categories of expertise in this stage, how can we ensure that the expertise of our community partners informs these inquiries? And, finally, what concrete steps can we take to address these structural barriers to sustain our interventions? The deliverables produced here should include the formation of robust teams of experts as well as specific steps that can be taken to deploy this collective knowledge to address the social and policy environments that create the need for our innovative interventions and in which our patients and community partners live.

Sociotechnical Tools to Address Environmental Barriers to Health

Our interview participants and community design team members in Flint, a predominantly Black and low-income city that is recovering from an economic downfall following the departure of the General Motors [20] automotive plant as well as an ongoing toxic water crisis [21,22], report that their capability and motivation to engage in physical activity to lower their blood pressure is often limited by their opportunity to do so. They identify 2 persistent barriers here: a dearth of safe outdoor spaces for exercise such as parks and sidewalks on which to traverse their neighborhoods, and the high cost of gyms and other indoor spaces where they might use exercise equipment for sustained physical activity. Even when park space was available, as one participant, a 61-year-old Black man told us, "I don't see anything for seniors." This absence of available public space made the challenge of affording a gym membership, even before the ongoing COVID-19 pandemic contributed to massive economic insecurity, more difficult. As another participant, a 52-year-old Black woman told us, "I would love to be able to have a gym membership. But there's only a certain amount of income—I'm on a fixed income...But everything is so expensive, it's difficult."

Notably, neither a lack of access to spaces for physical activity nor limited financial resources presented as barriers to capability or opportunity for our participants in Ann Arbor, the affluent city in which the University of Michigan campus where we work is located. As another participant, a 48-year-old Asian woman, remarked, "There's a lot of paths for walking, and we live near a playground so you can do something with the playground...They have equipment for you." Likewise, income did not emerge as a barrier to physical activity within this population. Another participant, a 72-year-old White woman, said, "We're fortunate we are retired, we have income, we aren't dependent on the job anymore, we have a pension and social

security. We have social security in the literal sense of secured money, and being able to afford a gym membership or any equipment."

These are thorny issues involving complex interactions between race, class, income, geography, and public policy; if we seek to understand how technology can be used to promote healthy behaviors among patients, we must begin by acknowledging that these technologies are sociotechnical tools that, by definition, emerge from the interaction of these social forces. These technologies do not exist within a bubble and, as we argue here, neither can our efforts to design them.

The Importance of Directly Addressing Structural Barriers to Health in mHealth Design

What we confront when we do health services research involving digital health tools is the fundamental tension between the promise of these devices and the ethos of innovation that spirits them, and the much more challenging realities of structural barriers to health that enable racial health disparities to persist. Melissa S. Creary [23] theorizes this tension through the concept of "bounded justice," a phenomenon where the good intentions of justice-oriented stakeholders "are bounded by greater socio-historical constraints." It is not enough, Creary writes, to pursue health equity and the amelioration of the indignities of longstanding health inequities simply through the distribution of "goods, materials, and resources" [23]. The political idealism of such interventions, even among so-called justice-based inclusive programs, comes with inherent limitations in its ability to repair "the underlying and deeper social inequalities embedded in individuals and communities, specifically those disadvantaged by racism," when they fail to address "the underlying mechanisms that generated initial historical inequalities" in the first place [23]. We echo Creary's call for more reflexive thinking about and action in the name of justice-oriented work that addresses these mechanisms as a part of our digital health technology intervention design.

Keeping these structural conditions at the forefront of our thinking about digital health design is vital because, lest we forget, we are not designing technical fixes to disease disparities but, rather, sociotechnical ones that must also engage with the social and policy environments that both necessitate innovation and constrain its deployment. This is why we argue early in this article that techno-pessimism and an effort to continue to improve digital health technologies constitute a compatible position and a requisite relationship with technology. The structural conditions that inhibit the sustained use of the tools we hope will help improve patient health should not dissuade us from seeking to improve care; rather, they should drive us to think more expansively, ethically, and systematically about this work. They should motivate us to foster collaborative relationships with policy experts, media and informatics scholars, and historians of medicine, as well as patients and caregivers with a wide range of interests in and objections to these kinds of technologies. They should center the role of structural racism in limiting access to lifesaving resources and in reproducing health disparities. And they should highlight opportunities for providers and researchers to contribute to the existing work that patient communities are engaging in to undo these structural harms.

Beginning from an acknowledgement that we are addressing "deep social problems" [24], which our technological interventions are simply unable to solve, enables us to identify social policy approaches that may help providers and patients to make the long-term improvements to health that they seek. This work must begin with an assessment of how we think about the role of technology in our research. As we are reminded by Amy Moran-Thomas's [25] writing about the use of the pulse oximeter during the ongoing pandemic, we must be self-reflexive and critical of the tools we design and deploy to identify how our technologies might reproduce racial health disparities. And we must acknowledge that our focus on individual-level health behaviors, as Veinot et al [8] warn us, can only get us so far.

Conclusions

Reframing how we think about our work—so that these issues and local contexts closely inform how we define our research problems, the kinds of solutions we pursue, and the changes we work to develop—can help us to bridge the divide between the techno-optimist and techno-pessimist positions in digital health research. We should be driven by this concern about "intervention-generated inequalities" to engage critically and productively with the promise of mHealth, ensuring that our work addresses the systems through which health disparities persist. We should think about the policy questions our data can illuminate, and make coalition building within and outside of our traditional networks of expertise an essential part of our work [2,26]. And when our work, by definition, involves the development and deployment of digital health technologies in an effort to improve health outcomes, we should integrate critical reflection about the technologies we deploy, the social contexts in which our patients and community partners live, and concerns about structural barriers to health as part of our efforts to design just and equitable health interventions.

Acknowledgments

The authors acknowledge their community partners in Ann Arbor and Flint, Michigan, for their time and work in our design process.

This work was supported by an award from the American Heart Association (20-SFRN35370008, Wearables in Reducing risk and Enhancing Daily Lifestyle [WIRED-L]).

Conflicts of Interest

None declared.

References

1. Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, American Heart Association Council on EpidemiologyPrevention Statistics CommitteeStroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2020 Update: A Report From the American Heart Association. Circulation 2020 Mar 03;141(9):e139-e596 [FREE Full text] [doi: 10.1161/CIR.0000000000000757] [Medline: 31992061]
2. Skolarus LE, Sharrief A, Gardener H, Jenkins C, Boden-Albala B. Considerations in Addressing Social Determinants of Health to Reduce Racial/Ethnic Disparities in Stroke Outcomes in the United States. Stroke 2020 Nov;51(11):3433-3439 [FREE Full text] [doi: 10.1161/STROKEAHA.120.030426] [Medline: 33104471]
3. Heron M. Deaths: Leading Causes for 2015. Natl Vital Stat Rep 2017 Nov;66(5):1-76 [FREE Full text] [Medline: 29235984]
4. Fryar CD, Chen T, Li X. Prevalence of uncontrolled risk factors for cardiovascular disease: United States, 1999-2010. NCHS Data Brief 2012 Aug(103):1-8 [FREE Full text] [Medline: 23101933]
5. Anderson-Lewis C, Darville G, Mercado RE, Howell S, Di Maggio S. mHealth Technology Use and Implications in Historically Underserved and Minority Populations in the United States: Systematic Literature Review. JMIR Mhealth Uhealth 2018 Jun 18;6(6):e128 [FREE Full text] [doi: 10.2196/mhealth.8383] [Medline: 29914860]
6. The Lancet Digital Health. Every breath you take, every move you make. Lancet Digit Health 2020 Dec;2(12):e629 [FREE Full text] [doi: 10.1016/S2589-7500(20)30272-7] [Medline: 33328027]
7. Toyama K. Technology as Amplifier in International Development. 2011 Presented at: iConference; Feb 8-11, 2011; Seattle, WA URL: https://tinyurl.com/5n6sw447 [doi: 10.1145/1940761.1940772]
8. Veinot TC, Ancker JS, Cole-Lewis H, Mynatt ED, Parker AG, Siek KA, et al. Leveling Up: On the Potential of Upstream Health Informatics Interventions to Enhance Health Equity. Med Care 2019 Jun;57 Suppl 6 Suppl 2:S108-S114. [doi: 10.1097/MLR.0000000000001032] [Medline: 31095048]
9. Schüll ND. Data for life: Wearable technology and the design of self-care. BioSocieties 2016 Oct 13;11(3):317-333. [doi: 10.1057/biosoc.2015.47]
10. Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, et al. Mobile health technology evaluation: the mHealth evidence workshop. Am J Prev Med 2013 Aug;45(2):228-236 [FREE Full text] [doi: 10.1016/j.amepre.2013.03.017] [Medline: 23867031]
11. Martin T. Assessing mHealth: opportunities and barriers to patient engagement. J Health Care Poor Underserved 2012 Aug;23(3):935-941. [doi: 10.1353/hpu.2012.0087] [Medline: 24212144]
12. Lupton D. The digitally engaged patient: Self-monitoring and self-care in the digital health era. Soc Theory Health 2013 Jun 19;11(3):256-270. [doi: 10.1057/sth.2013.10]
13. Anderson M. Mobile Technology and Broadband. Pew Research Center. 2019. URL: https://www.pewresearch.org/internet/2019/06/13/mobile-technology-and-home-broadband-2019/ [accessed 2022-05-27]
14. Peek ME. Can mHealth Interventions Reduce Health Disparities among Vulnerable Populations? Divers Equal Health Care 2017;14(2):44-45. [doi: 10.21767/2049-5471.100091]
15. Krishnaswami A, Beavers C, Dorsch MP, Dodson JA, Masterson Creber R, Kitsiou S, Innovations, Cardiovascular Teamthe Geriatric Cardiology Councils, American College of Cardiology. Gerotechnology for Older Adults With Cardiovascular Diseases: JACC State-of-the-Art Review. J Am Coll Cardiol 2020 Dec 01;76(22):2650-2670 [FREE Full text] [doi: 10.1016/j.jacc.2020.09.606] [Medline: 33243384]
16. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
17. Flint Community. Healthy Flint Research Coordinating Center. URL: https://www.hfrcc.org/resources [accessed 2022-05-30]
18. Gilliard C, Culik H. Digital Redlining, Access, and Privacy. Common Sense. 2016. URL: https://www.commonsense.org/education/articles/digital-redlining-access-and-privacy [accessed 2022-05-27]
19. Merid B, Robles MC, Nallamothu BK. Digital Redlining and Cardiovascular Innovation. Circulation 2021 Sep 21;144(12):913-915. [doi: 10.1161/circulationaha.121.056532]
20. Highsmith AR. Demolition Means Progress: Flint, Michigan, and the Fate of the American Metropolis. The University of Chicago Press. ISBN. Chicago, IL: The University of Chicago Press; 2015.
21. Masten SJ, Davies SH, Mcelmurry SP. Flint Water Crisis: What Happened and Why? J Am Water Works Assoc 2016 Dec 01;108(12):22-34 [FREE Full text] [doi: 10.5942/jawwa.2016.108.0195] [Medline: 28316336]
22. Hanna-Attisha M, LaChance J, Sadler RC, Champney Schnepp A. Elevated Blood Lead Levels in Children Associated With the Flint Drinking Water Crisis: A Spatial Analysis of Risk and Public Health Response. Am J Public Health 2016 Feb;106(2):283-290. [doi: 10.2105/AJPH.2015.303003] [Medline: 26691115]
23. Creary MS. Bounded Justice and the Limits of Health Equity. J Law Med Ethics 2021 Jun 29;49(2):241-256 [FREE Full text] [doi: 10.1017/jme.2021.34] [Medline: 34924041]

24. Glabau D. The Dark Matter of Digital Health. Public Books. 2020. URL: https://www.publicbooks.org/the-dark-matter-of-digital-health/ [accessed 2022-05-27]
25. Moran-Thomas A. How a Popular Medical Device Encodes Racial Bias. Boston Review. 2020. URL: http://bostonreview.net/science-nature-race/amy-moran-thomas-how-popular-medical-device-encodes-racial-bias [accessed 2022-05-27]
26. Merid B, Whitfield CO, Skolarus LE. Reflections on the Values of Community-Based Participatory Research in Supporting Mobile Health Technology Use. Circ Cardiovasc Qual Outcomes 2020 Sep;13(9):e007131 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.120.007131] [Medline: 32885680]

Abbreviations

CERB: Community Ethics Review Board
COM-B: Capability, Opportunity, Motivation, Behavior
mHealth: mobile health
WIRED-L: Wearables in Reducing Risk and Enhancing Daily Lifestyle

Edited by L Buis; submitted 08.06.21; peer-reviewed by Y Li, R Marshall, D Frost; comments to author 27.10.21; revised version received 31.03.22; accepted 29.04.22; published 10.06.22.

Please cite as:
Merid B, Robles MC, Nallamothu BK, Newman MW, Skolarus LE
Viewing Mobile Health Technology Design Through the Lens of Amplification Theory
JMIR Mhealth Uhealth 2022;10(6):e31069
URL: https://mhealth.jmir.org/2022/6/e31069
doi:10.2196/31069
PMID:35687411

Viewpoint

Beyond Pathogen Filtration: Possibility of Smart Masks as Wearable Devices for Personal and Group Health and Safety Management

Peter Lee[1*], PhD; Heepyung Kim[1*], PhD; Yongshin Kim[2], BA; Woohyeok Choi[3], PhD; M Sami Zitouni[4,5], PhD; Ahsan Khandoker[4,5], PhD; Herbert F Jelinek[4,5], PhD; Leontios Hadjileontiadis[4,5,6], PhD; Uichin Lee[1,7], PhD; Yong Jeong[1,8], MD, PhD

[1]KAIST Institute for Health Science and Technology, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea
[2]Graduate School of Data Science, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea
[3]Information & Electronics Research Institute, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea
[4]Department of Biomedical Engineering, Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates
[5]Healthcare Engineering Innovation Center, Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates
[6]Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece
[7]School of Computing, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea
[8]Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea
[*]these authors contributed equally

Corresponding Author:
Yong Jeong, MD, PhD
Department of Bio and Brain Engineering
Korea Advanced Institute of Science and Technology
291 Daehak-ro Yuseong gu
Daejeon, 34141
Republic of Korea
Phone: 82 423507165
Email: yong@kaist.ac.kr

Abstract

Face masks are an important way to combat the COVID-19 pandemic. However, the prolonged pandemic has revealed confounding problems with the current face masks, including not only the spread of the disease but also concurrent psychological, social, and economic complications. As face masks have been worn for a long time, people have been interested in expanding the purpose of masks from protection to comfort and health, leading to the release of various “smart” mask products around the world. To envision how the smart masks will be extended, this paper reviewed 25 smart masks (12 from commercial products and 13 from academic prototypes) that emerged after the pandemic. While most smart masks presented in the market focus on resolving problems with user breathing discomfort, which arise from prolonged use, academic prototypes were designed for not only sensing COVID-19 but also general health monitoring aspects. Further, we investigated several specific sensors that can be incorporated into the mask for expanding biophysical features. On a larger scale, we discussed the architecture and possible applications with the help of connected smart masks. Namely, beyond a personal sensing application, a group or community sensing application may share an aggregate version of information with the broader population. In addition, this kind of collaborative sensing will also address the challenges of individual sensing, such as reliability and coverage. Lastly, we identified possible service application fields and further considerations for actual use. Along with daily-life health monitoring, smart masks may function as a general respiratory health tool for sports training, in an emergency room or ambulatory setting, as protection for industry workers and firefighters, and for soldier safety and survivability. For further considerations, we investigated design aspects in terms of sensor reliability and reproducibility, ergonomic design for user acceptance, and privacy-aware data-handling. Overall, we aim to explore new possibilities by examining the latest research, sensor technologies, and application platform perspectives for smart masks as one of the promising wearable devices. By integrating biomarkers of respiration symptoms, a smart mask can be a truly cutting-edge device that expands further knowledge on health monitoring to reach the next level of wearables.

(JMIR Mhealth Uhealth 2022;10(6):e38614) doi:10.2196/38614

KEYWORDS

smart mask; pathogen filtration; COVID-19; protective equipment; digital health; wearable; smart device; wearable device; sensor; health monitoring

Introduction

After the World Health Organization declared COVID-19 a "pandemic" (a global epidemic) attributed to SARS-CoV-2 infection [1], masks have been used by the general population all over the world for precautionary health reasons [2,3]. As a result, people wear masks at all times and in all places; however, the pandemic has revealed the limitations of current mask deployments regarding not only the spread of the disease but also concurrent psychological, social, and economic complications.

To improve these limitations, smart face masks designed with electronic sensors have been recently proposed. The continuous use of masks has led to the designs of various face mask products, which have become available on the market. The term "smart" has been used to signify possible additional functionalities of the "smart (face) masks" around the world, leading to an expansion of the mask's usage, including masks for protection, health, and environmental sensing [4-6].

While the COVID-19 pandemic is seemingly under control owing to vaccination, there is a need for innovative, Internet of Things (IoT)–based smart-mask solutions to help people transition to a postpandemic world, where the emergence of infectious SARS-CoV-2 variants is prevalent along with the heightened possibility of further, yet unknown, virus pandemics, and to combat airborne diseases [7,8]. In combination with data-driven applications, IoT and smart connected technologies can play a critical role in individual protection and extend to group sensing for the prevention, mitigation, and continuous remote monitoring of patients. Such a benefit of group sensing is shown with a contact-tracing app, where it could instruct a person in close contact with patients with COVID-19 to quickly self-isolate to reduce disease transmission [9].

Here we present a viewpoint for smart masks in the form of emerging IoT-based solutions by examining the current status of smart masks, potential sensors for their functional expansion, connected architecture of smart masks for individual and group health care, and further considerations for actual deployment of such technology in the field. The details are as follows:

- Current status of existing commercial and academic smart masks
- Smart mask expansion in terms of personal health care and disease diagnosis
- Connected architecture and applications of smart masks
- Further real-world considerations

Features and Applications of Current Smart Masks in the Field

Relevant smart masks available in the market were found through web searches, including Amazon, using the following search terms: "Smart Mask," "Facial," and "Electronics." The search for publications was performed using 5 databases (Google Scholar, Web of Science, ScienceDirect, PubMed, and EBSCO) on the basis of the following combinations of search terms: "Smart mask" OR "Smart face mask," "sensor," "IoT," AND "Healthcare."

We defined 3 major inclusion criteria of reports on smart masks in this review.. Specifically, these criteria involve the following: (1) sensing: sensors attached to the mask; (2) actuation: functional manipulation of the mask; and (3) connectivity: communicating sensor data using mobile, cloud storage, or IoT-based networks. Only articles published between January 2020 and May 2022 were included to examine smart masks developed after the COVID-19 outbreak. Finally, the study selection procedure resulted in 12 smart mask products and 13 smart mask research prototypes reported in this study. Tables 1 and 2 list their functions and features, respectively. Detailed selection criteria are provided in Multimedia Appendix 1.

Table 1. Commercially available smart masks with their key features.

Name and purpose	Function	Feature
Air control with respiration rate–sensing		
AO AIR Atmos mask [10]	Automatic fan control with respiration rate–sensing and filter status check	• S[a]: Filter status and respiration • A[b]: Fan on/off control • C[c]: Bluetooth
LG PuriCare (2nd Gen) [11]	Automatic fan control with respiration rate–sensing	• S: Respiration rate • A: Fan on/off control • C: Bluetooth
Ventilation		
ATMOBLUE Face Mask [12]	Three fan speed modes and air quality check	• S: Air quality • A: Fan speed control • C: Bluetooth
Belovedone Air Purifier [13]	Two fan speed modes	• A: Control fan speed
Philips Fresh Air Mask [14]	Three fan speed modes	• A: Control fan speed
Xiaomi Purely [15]	Three fan speed modes	• A: Control fan speed
CSE&L AIRVISOR [16]	Three fan speed modes	• A: Control fan speed
CELLRETURN CX9 [17]	Sterilization and LED skin care	• A: LED[d] sterilization and skin care
Razer Zephyr [18]	Two fan speed modes and lighting	• A: Control fan speed and customizable lighting zones • C: Bluetooth
Communication aid		
CLIU Pro [19]	Air quality check and built-in microphone	• S: Air quality, mask wear time, and head motion • C: Bluetooth
Donut Robotics C-FACE [20]	Speech to text and voice translation	• A: speech-to-text message, voice call, and translation • C: Bluetooth
TrendyNow365 LED Mask [21]	Text display on mask surface	• A: Display custom LED letters • C: Bluetooth

[a]S: sensing.

[b]A: actuation.

[c]C: connectivity.

[d]LED: light-emitting diode.

Table 2. Smart mask research prototypes from academic journals.

Name and purpose	Function	Feature
External pathogen detection and elimination		
ADAPT [22]	Pathogen sensing and mist spray activation	• S[a]: Airborne particle sensing • A[b]: Mitigation module on/off • C[c]: Bluetooth
COVID-19 detection		
SARS-CoV-2-sensing face mask [23]	Detects COVID-19 infection	• S: Paper-based nucleic acid diagnostics
Lightweight and zero-power smart face mask [24]	Monitor cough and check mask-wearing	• S: Mask deformation • C: RF[d] transponder
AG47-SmartMask [25]	Monitor cardio-respiratory variables and to detect cough	• S: Breathe pattern, skin/DSV[e] temperature, humidity, air pressure, HR[f], and SpO2[g] • C: Bluetooth
Respiratory disease–monitoring		
Smart face mask with Heat flux sensor [26]	Noninvasive body temperature and breathing rate–monitoring	• S: Facial skin temperature and breathing rate • C: LoRa[h] and Wi-Fi
Smart facemask for wireless CO_2 monitoring [27]	Monitor CO_2 in DSV	• S: CO2 concentration • C: NFC[i]
Smart face mask with ultrathin pressure sensor [28]	Breath monitoring	• S: DSV pressure change • C: Wi-Fi connection
Smart face mask with wearable pressure sensor [29]	Breath monitoring	• S: DSV pressure change • C: Bluetooth connection
Smart medical mask for health care personnel [30]	Detect respiratory breathing, fever, and alert possible face irritation	• S: DSV temperature, mask strain • C: Wi-Fi
Lab-on-Mask [31]	Monitor cardio-respiratory variables	• S: HR, BP[j], SpO2, and skin temperature • C: Bluetooth connection
General health monitoring		
FaceMask [32]	Monitor cardio-respiratory variables and mask-wearing	• S: Humidity, DSV or external temperature, volatile organic compounds. And head motion • C: Bluetooth connection
Facebit [33]	Monitor HR, respiration rate, mask fit, and wear time	• S: HR, respiration rate, mask fit, and wear time • C: Bluetooth
Masquare [34]	Monitor cardio-respiratory variables	• S: Respiratory pressure, HR, SpO2, and head motion • C: Bluetooth

[a]S: sensing.

[b]A: actuation.

[c]C: connectivity.

[d]RF: radiofrequency.

[e]DSV: dead space volume.

[f]HR: heart rate.

[g]SpO_2: blood oxygen saturation.

[h]LoRa: long range.

[i]NFC: near-field connection.

[j]BP: blood pressure.

Most commercial masks used in daily life provide actuations based on use, such as exchangeable filters, self-sterilizers, embodied microphones, and integrated fans. In total, 4 smart masks had sensing capabilities such as air pathogen check, filter status, and breath monitoring. In total, 11 smart masks included actuation with mostly inner fan speed control and LED lighting control. A total of 7 smart masks supported a connectivity feature through a Bluetooth connection with the smartphone. The masks that supported all 3 features (ie, sensing, actuation, and connectivity) were those of Atmos AO AIR [10], LG PuriCare (2nd Gen) [11], and ATMOBLUE [12]. These smart masks offer inner fan control actuation and Bluetooth connectivity while using different sensing (filter, respiration rate, and air-quality checks). Commercial masks have focused on mitigating discomfort such as breathing difficulty, excessive moisture inside the mask, fogging of glasses, and hygiene problems caused by long-term use [35-37]. Besides protection, the masks of CLIU [19], Donut Robotics [20], and TrendyNow365 [21] aimed to overcome speech problems with mask-wearing. Additional investigations, such as mask material, weight, and battery usage time, are presented in Multimedia Appendix 1.

While commercial smart masks were focused on user comfort, academic prototypes were designed for sensing capabilities such as health monitoring and disease detection. For example, in terms of COVID-19 detection, Nguyen et al [23] integrated a cell-free sensor to detect SARS-CoV-2, and Ye et al [24] and Fois et al [25] focused on detecting abnormalities such as coughing behavior. Not specific to COVID-19 but to cope with general respiratory disease, Lazaro et al [26], Escobedo et al [27], Zhong et al [28], Yang et al [29], Kim et al [30], and Pan et al [31] monitored breathing patterns. From a general health monitoring perspective, Gravina et al [32], Curtiss et al [33], and Fischer et al [34] monitored biosignals such as heart rate, respiration rate, and body temperature. Acquired sensor readings were then analyzed through smartphone apps for display.

All prototype masks were considered with regard to their physiological sensing capabilities. A total of 12 smart masks were considered with connectivity features using Bluetooth connectivity, near-field communication (NFC), a long range, and Wi-Fi connectivity with the smartphone. Ye et al [24] further demonstrated a radiofrequency (RF) feature using silver nanowires attached to the inner layer for monitoring cough and mask usage. Overall, the current features of smart masks available in the market offer environmental (air quality) monitoring, mask quality–monitoring, and functions for user comfort. On the other hand, research prototypes can be summarized as health monitoring and respiratory disease detection.

Possible Directions for Feature Extension

Our investigation of research prototypes showed that existing masks support health monitoring and disease diagnosis on the basis of vital signs such as respiration, blood oxygen saturation, and body temperature. In this section, we further explore what other biosignals can be measured and what applications can be used through a smart mask as a wearable device for health care and safety. In addition, we argue that it is critical to reduce the posterior auricular (back of the ear) discomfort and pain caused by long-term wearing of the mask, as witnessed by a mask frame extension that supports an ear strap introduced recently [38]. In consideration of the ear strap frame, we would like to present a viewpoint on the extension of the application of the smart mask and its potential as a biosignal measuring device. To systematically search for feasible sensors, the expressions “smart” and “intelligent” textiles or “wearable electronic” are keywords used for selection. Sensors that sense and react to biosignals, environmental conditions, or stimuli, such as those from breath, skin, head motion, air, or other sources, were investigated. Multiple biosignal sources can be recorded around the face with sensors incorporated into the smart masks to measure biosignals and interior or exterior environmental factors [22,39].

For the facial part of the mask, pressure sensors can be used to obtain the respiration rate and inhalation volume to monitor breathing patterns [28,29]. These are piezoelectric-like sensors that are sensitive enough to respond to exhale volume pressure and flexible, lightweight, and energy-efficient circuits that can fit into the mask. With continuous monitoring of breathing patterns, we expect to observe users’ lung health or screen patients with chronic lung disease [40]. In addition to analyzing breath, chemical sensors can be used as markers for personal health problems and respiratory diseases by targeting specific molecules [41-48]. These sensors are based on metal oxides whose target compounds can be easily switched with specific reagents. Several applications include acetone for diabetes [41-43], hydrogen sulfide for small intestinal bacterial overgrowth [44,45], and toluene for lung cancer diagnosis [46-48].

As the mask directly contacts the facial skin, a photoplethysmographic (PPG) sensor can be adopted to conduct pulse oximetry and measure heart rate variability, oxygen saturation, and blood pressure. These metrics are widely researched for indirect measures of physical and mental health [49-51], physical stress [31,49], and hypertension or hypotension [51-53], respectively. Electrooculography (EOG) [54,55], electrodermal activity (EDA) [56,57], and electromyography (EMG) [50,52,53] can also be adopted to measure various biophysical signals that arise from facial skin. For example, eye-blinking EOG measures have been linked to attention [58] and may infer the user’s mental state. The electrodermal response from EDA and facial muscle activation from EMG can be used as a measure of emotion such as anxiety or depression [55-57,59]. In addition, facial surface EMG was adopted for monitoring pain through facial expressions [60] (Table 3).

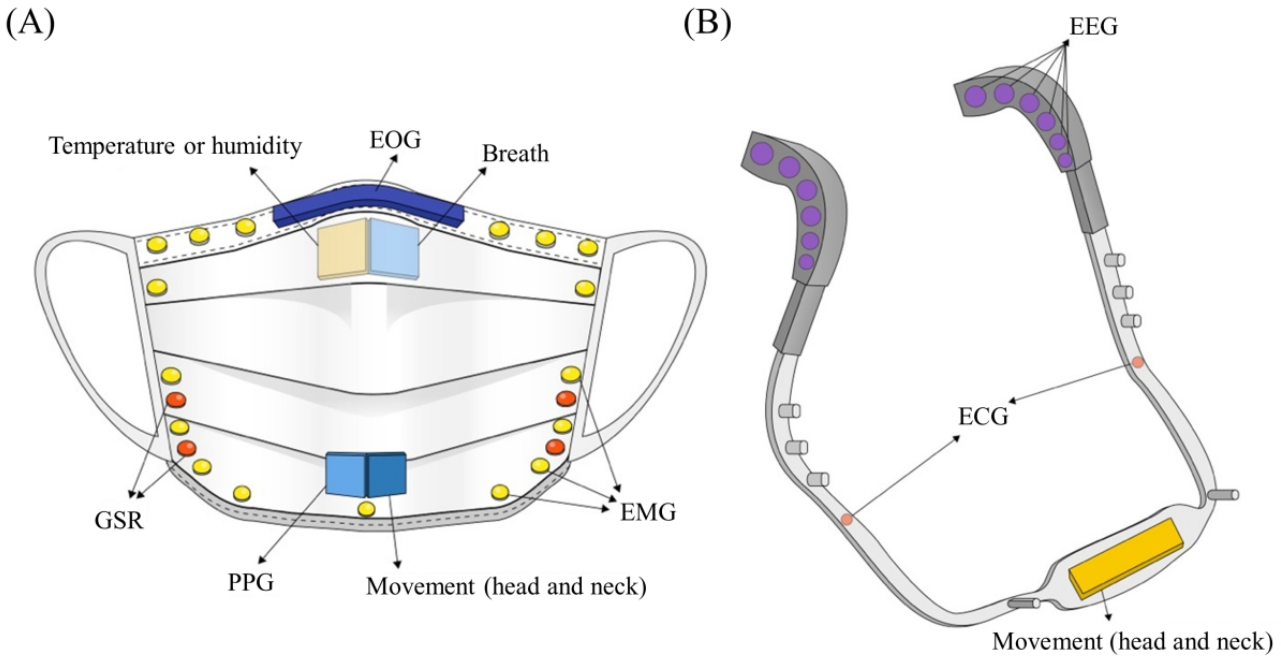
Table 3. Possible sensor integration on the masks.

Sensors and features	Applications
Location: mask main body	
Type: biosignal information	
Source: breath (respiration)	
Pressure sensor	
Respiration rate or volume	Personal health or sport [27,28]
Chemical sensor	
Ketone: acetone	Personal health or disease (diabetes) [41-43]
Hydrogen sulfide	Personal health [44,45]
Toluene	Personal health or disease (lung cancer) [46-48]
Source: facial blood vessels	
Photoplethysmography sensor	
Heart rate variability	Physical health or mental health [49-51]
Oxygen saturation	Physical stress [31,49]
Blood pressure	Hypertension or hypotension [51-53]
Source: skin	
Electrooculography sensor	
Eye blink	Concentration [54,55]
Electrodermal activity sensor	
Electrodermal response	Emotion [56,57]
Temperature sensor	
Temperature change	Communicable diseases [25,26,31,32]
Electromyography sensor	
Facial muscle	Emotion [55,59]
Facial muscle	Pain [60]
Source: head	
Inertial measurement unit	
Motion	Posture [61]
Type: environmental information	
Source: air	
Chemical sensor	
Environment air quality	Local air quality [22,39]
Source: external temperature	
Thermometer	
Temperature	Local temperature [62]
Source: external humidity	
Humidity sensor	
Humidity	Local humidity [63-65]
Location: mask support frame	
Type: biosignal information	
Source: ear	
Electroencephalography sensor	
Brain activity	Drowsiness or fatigue [66]

Sensors and features	Applications
Source: neck	
Inertial measurement unit sensor	
Motion	Posture [61]
Electrocardiographic sensor	
Heart	Heart disease [67,68]

A smart mask can also measure air pollution and several other environmental variables such as air quality [22,39], temperature [62], and humidity [63-65]. The inclusion of sensing air-tightness and the quality of filters can help ensure the additional benefits of smart masks by improving safety by providing an air-tight fit around the face. If the mask uses a support frame, such as a head or neck strap, electroencephalography (EEG) and electrocardiography (ECG) sensors can be applied to measure the electrical activity of the brain and heart. EEG signals have been used to detect a user's fatigue or drowsiness like fatigue in driving [66]. Integrating ECG can be an advantage over PPG readings as it records the heart's electrical activity at its source [67,68]. Lastly, inertial measurement unit sensors can be attached to the ear strap for activity sensing that can discern fall or head collision [61]. The possible sensor attachments on a facial mask and ear strap are depicted in Figure 1.

Figure 1. The possible sensor attachments on (A) a facial mask and (B) the ear strap. ECG: electrocardiography; EEG: electroencephalography; EMG: electromyography; EOG: electrooculography; GSR: Galvanic skin response; PPG: photoplethysmography.



Toward Connected Smart Masks

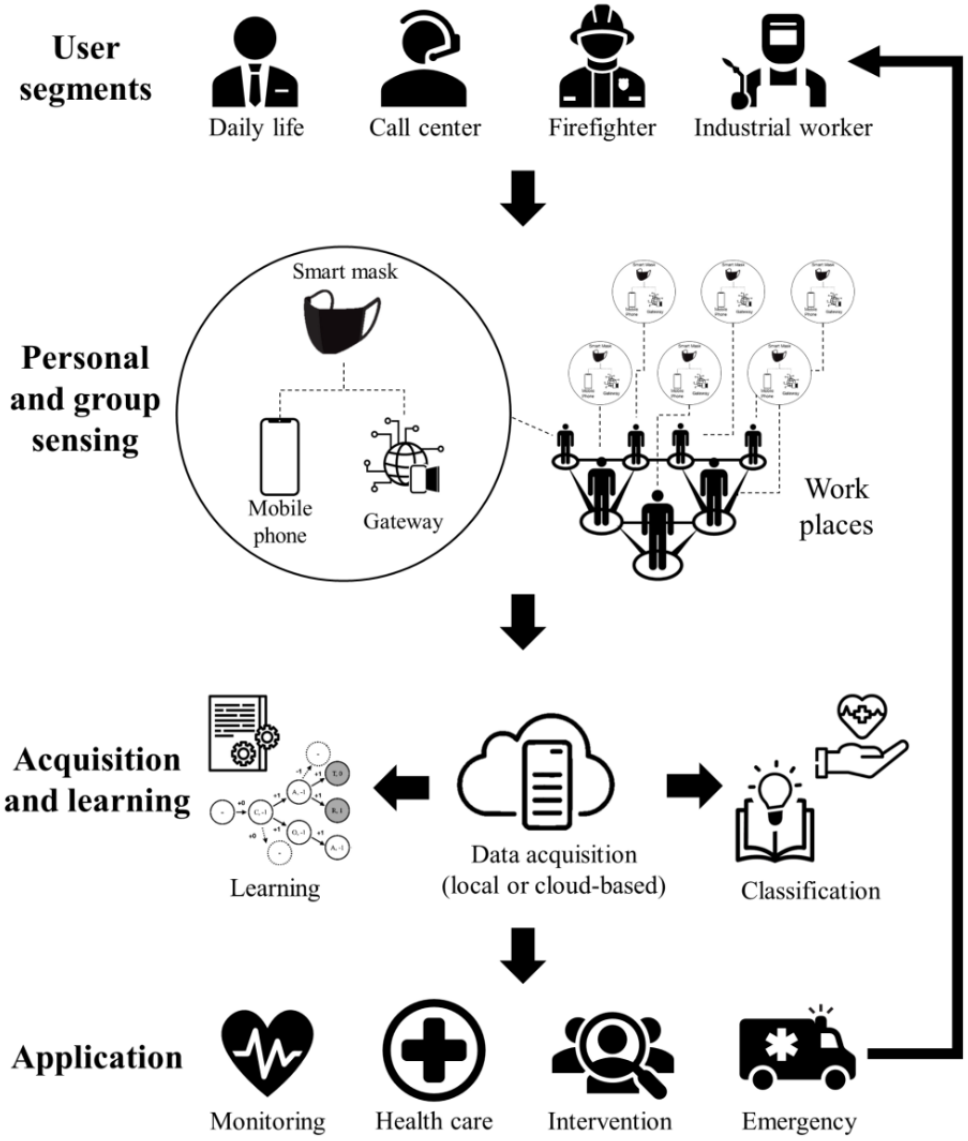
In this section, we attempt to seek opportunities beyond personal protective equipment to group management, so-called group-sensing, through connected smart masks as wearable devices for health care and safety. The advantages of group-sensing include continuously measuring and managing a population's physical and mental health through the sensors inside the smart mask or via connected smart mask platforms. Such advantages are particularly useful in dealing with infectious diseases that spread through contact and saliva, such as COVID-19 [69]. The smart masks of those at risk can be managed, and remote caregiving can be supported via connected devices. As in a prior study on smartwatches [70], their everyday health conditions (eg, breathing and heart rates) can be tracked and analyzed to detect early signs of respiratory behavior changes, which could be related to COVID-19 infection. Namely, beyond a personal sensing application, a group or community sensing application may share an aggregate version of information with the broader population. The architecture of connected masks is shown in Figure 2 by extending prior mobile sensing architecture [71,72].

For group sensing, the smart mask should be able to transmit the collected data to the server by using wireless communication protocols such as Wi-Fi, long-term evolution, 4G and 5G networks, Zigbee, and narrowband IoT without manual operation [73]. Besides, the analysis results should allow the user to take action or receive an alarm related to a particular hazard. Most smart masks integrate communication modules to use smartphones for displaying sensing results and as a gateway terminal to interact on the web [27,33,74,75]. In addition, smartphones allow short-distance connections such as Bluetooth,

NFC, and radiofrequency identification, where acquired data can be transferred to local IoT gateways [73]. Furthermore, server clouds and relevant analytics technology are required to store smart mask data and process large sets of data to develop applications such as health care, safety monitoring, and intervention for the users. This kind of collaborative sensing will also address the challenges of individual sensing, such as reliability and coverage [76].

The information gathered in cloud servers can be used with machine learning (ML) and data mining applications [75,77]. The advantages of utilizing ML for group sensing results are system optimization and acquired data processing [77]. For instance, collecting data on device failures, usage time, filter, and battery can be analyzed for design considerations and maintaining the optimal operation of a smart mask. Furthermore, Gravina et al [32] reported the application of ML in smart masks, where they tested mask wear classification from sensor signals. In terms of data mining, a more detailed air quality map can be created as the user wears a smart mask with environmental sensors and moves around places collecting data. Moreover, GPS for community sensing can facilitate real-time sensing and location-based monitoring of masks and actions of multiple users in some local environments, such as COVID-19 contact-tracing, local airborne pathogen detection, or emergency services.

Figure 2. Connected smart mask architecture.



With modern technological advances, it has become possible to collect big data and create new knowledge that we have not been able to analyze before. Unlike conventional wearable devices, smart masks can collect biomarkers of respiration or the respiratory system and expand further knowledge on wearables. Previous work by Curtiss et al [33] and Hyysalo et al [75] shows detailed aspects of the connected smart mask platform and deployment considerations. Curtiss et al's [33] Facebit smart mask accompanies a mobile app that displays sensing results such as heart rate, respiration rate, mask fit, and wear time. This app communicates with Facebit through Bluetooth and stores data in a local database. For now, stored data are used to track a user's mask-wearing time and send a notification to replace the mask. As an open-source smart mask

research platform, this work demonstrates proof-of-concept connected smart masks and presents further research on personalizing algorithms and applications for respiratory health tools. Hyysalo [75] illustrated the software architecture of the smart mask platform, including the mask, mobile app, and backend health artificial intelligence. In addition, this study envisioned a smart mask ecosystem [78,79]—a collection of infrastructure, analytics, and applications, to draw personal health trajectories.

Further Considerations for Real-World Use

Lastly, we present and discuss viewpoints on the application fields of the connected mask and further considerations for practical use. As the smart face mask is a promising respiratory monitoring tool, we explored relevant fields where it can benefit direct needs. Aside from the primary field of daily-life health-monitoring, we envision several real-world uses such as sports training, ambulatory setting, industry and firefighter safety masks, and military applications. In the following sections, several directions for real-world deployment scenarios of smart masks are first discussed. Thereafter, we discuss sensor accuracy and reproducibility issues, most critical ones in measuring biosignals through all wearable devices. Ergonomic design for the general population needs to be considered for public acceptance of smart masks. Finally, privacy-aware data-handling is necessary for security to collect and manage personal biosignals.

Service Application for Real-World Use

Daily Life Health Monitoring

The smart mask presents an opportunity to apply advanced analytics to health care. The analysis of physiological changes, such as breathing pattern, pulse rate, and tidal volume, enables us to monitor respiratory health, diagnose relevant diseases, and point of care through continuous monitoring. In addition, other various features can be obtained, as we discussed in the possible sensor extension scenarios, for instance, stress and fatigue [80].

Sports Training

In particular, smart masks can be adopted for measuring the cardiopulmonary exercise load, which is an important index in evaluating exercise capacity. Previously, this was done by wearing additional equipment in wired or wireless form with controlled settings [81]. This test can be easily accessible to the general population; for example, in a gymnasium or through home-based training through smart mask application. Furthermore, owing to the recent COVID-19 pandemic, there is increasing demand for indoor exercise platforms such as Zwift [82], where individuals can virtually compete with users on the internet and measure exercise ability and improvements. The smart mask can contribute as a wearable device for additional exercise measures in such settings.

Emergency Room or Ambulatory Settings

In the emergency room or ambulatory settings, masks have been used to deliver air and monitor respiration. We expect smart masks to be adopted to track health status without any additional device. Additionally, nosocomial infections, such as ventilator-associated pneumonia, can be detected with the use of the smart mask [83].

Industry Workers and Firefighters

Many workers at coal mines, construction sites, and chemical plants and firefighters at fire scenes are prone to hazardous gas; thus, wearing a mask is mandatory for safety issues. Smart masks can be used to track the health status of people who have been poisoned by gas or toxic substances or have been exposed by measuring the surrounding situation. Besides, real-time environmental monitoring can ensure user safety and prompt responses to fast-changing hazardous events through the detection of gas leakage or toxic events [84].

Soldier Safety and Survivability

Recently, there has been ongoing research on wearable devices such as vests and helmets to collect biosignals for the safety and survival of soldiers [85]. The smart mask can also be a promising wearable device in respiratory monitoring. It is expected that safety and survival can be further improved by collecting the soldier's biosignals, location information, or information about the surrounding environment. These measures help monitor the soldier's physical and mental health status and decision-making. Moreover, breath analysis can predict and monitor the onset of pulmonary injury due to various environmental and infectious exposures [86].

Accurate and Reliable Sensors

One major requirement for such predictive diagnostics is that sensor information must be accurate and reliable. The type of sensor and its placement affect the measurements. For instance, potential inaccuracies rise with excessive motion artifacts involving many physical activities, such as sports, firefighting, or military action. Although the reviewed articles described potential applications and demands for health intervention, they provided little evidence related to the usability and practicality of the proposed device. As the temperature and humidity rise owing to mask-wearing, the adhesion between the sensor and the skin may decrease, and sweat generated by humidity may negatively affect accurate sensor signal measurement. Beyond sensing accuracy and reliability, it is important to consider additional metrics, such as smart mask interoperability, versatility, power consumption, and durability, to examine the usefulness of the system as well as comfort and ease of use for different population characteristics [87,88].

Ergonomic Design for Usability

If users wear heavy equipment such as a helmet for a long time, it can strain their head and neck [89-91]. Masks with smart functions also increase in weight, unlike existing masks, owing to the addition of batteries, sensors, and fans. Therefore, it places a burden on the head and neck and may cause deformation in posture. If the systems within smart masks became more complicated, these could become more uncomfortable and make users reluctant to wear them. Detailed surveys on usability and performance evaluation from daily life trials need to be

conducted to ascertain the usability of smart masks [92-94]. Maximizing and optimizing the battery lifetime of the smart mask ensures user satisfaction and comfort [95,96]. If the device supports recharging, the rechargeable battery of the mask is a major contributor to the mask's weight. If the communication between the smart mask and the smartphone requires much energy and acquiring data from sensors may rapidly drain the battery, a larger battery capacity is then required. Thus, the overall weight of the mask increases. Therefore, in developing a smart mask, it is necessary to consider the battery size and material related to weight. In addition, since the material of their mask is in contact with the skin surface, it is necessary to use an approved suitable material [97]. Overall, the potential reluctance of users can be reduced by incorporating simple protocols for the number of sensors and user specificity, comfort, including weight, and fashion considerations for the general population [98,99].

Privacy-Aware Data-Handling

One challenge in developing connected smart mask architecture systems is the collection of personal information and privacy infringement. With the advancement of the IoT, real-time monitoring data are shared and analyzed to identify factors related to events. Although this monitoring is intended to assist users, some aspects of personal privacy are violated [100-104]. Prior studies have shown that privacy concerns related to wearable cameras are often influenced by users' social, behavioral, and environmental contexts [105]. For example, wearable camera users are often conscious of bystander privacy, and likewise, bystanders are concerned about potential privacy violations (eg, subtleness and ease of recording) [106]. In addition, advanced data processing methods may have privacy implications. For instance, personal physiological data or location information can be misused because of poor data management policies. In these scenarios, health monitoring results may encourage the tracking of work performance (ie, using the data for secondary purposes without explicit consent). This practice may influence the review of workers' performances and may cause monitoring to become a surveillance practice beyond health monitoring. Beyond secondary use, the security of the devices themselves can also be problematic, as the low computing power within smart mask systems may make them vulnerable to unauthenticated access [107,108]. As smart mask technology is still in its infancy, these implications are not yet fully understood and should be considered in future implementation strategies.

Conclusions

This study examined recent smart masks in conjunction with accompanying systems that could be used to prevent COVID-19 and other respiratory diseases. We then offered our viewpoints on smart masks in the form of emerging IoT solutions. Reviewing commercially available smart masks revealed the trend that smart masks were mainly designed to address user discomfort. However, recent research prototypes were taking further steps, not only dealing with COVID-19 but toward general health monitoring by supporting breathing and physiological signal sensing. Thus, we sought further functional expansion on smart masks by investigating previous mobile sensing studies. In addition, we extensively discussed novel opportunities for group health management through a connected smart masks platform. We believe that smart masks can serve as a truly cutting-edge device that expands the coverage of health monitoring and helps reach the next level of wearables.

Acknowledgments

This research was supported by the 2022 Smart project of the Korea Advanced Institute of Science & Technology and Khalifa University (KAIST-KU) Joint Research Center, KAIST, Daejeon, Korea, and the Basic Science Research Program through the National Research Foundation of Korea funded by the Korean government's Ministry of Science and Information and Communication Technology (2020R1A4A1018774).

Authors' Contributions

PL and HK wrote and equally contributed to drafting of the manuscript. YK performed a database search for study selection and wrote data for Tables 1 and 2. HK provided data for Table 3. WC, MSZ, AHK, HFJ, LH, UL, and YJ contributed to the critical revision of the paper, and all authors reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Table showing specifications of 12 commercially available smart masks.
[DOCX File , 19 KB - mhealth_v10i6e38614_app1.docx]

References

1. Coronavirus disease 2019 (COVID-19): situation report, 73. World Health Organization. URL: https://apps.who.int/iris/handle/10665/331686 [accessed 2022-06-10]

2. Howard J, Huang A, Li Z, Tufekci Z, Zdimal V, van der Westhuizen H, et al. An evidence review of face masks against COVID-19. Proc Natl Acad Sci U S A 2021 Jan 26;118(4):e2014564118 [FREE Full text] [doi: 10.1073/pnas.2014564118] [Medline: 33431650]
3. Cheng KK, Lam TH, Leung CC. Wearing face masks in the community during the COVID-19 pandemic: altruism and solidarity. The Lancet 2022 Apr;399(10336):e39-e40. [doi: 10.1016/s0140-6736(20)30918-1]
4. Lu L, Zhang J, Xie Y, Gao F, Xu S, Wu X, et al. Wearable Health Devices in Health Care: Narrative Systematic Review. JMIR Mhealth Uhealth 2020 Nov 09;8(11):e18907 [FREE Full text] [doi: 10.2196/18907] [Medline: 33164904]
5. Tian S, Yang W, Grange JML, Wang P, Huang W, Ye Z. Smart healthcare: making medical care more intelligent. Global Health Journal 2019 Sep;3(3):62-65. [doi: 10.1016/j.glohj.2019.07.001]
6. Perez AJ, Zeadally S. Recent Advances in Wearable Sensing Technologies. Sensors (Basel) 2021 Oct 14;21(20):6828 [FREE Full text] [doi: 10.3390/s21206828] [Medline: 34696040]
7. Zhou SS, Lukula S, Chiossone C, Nims RW, Suchmann DB, Ijaz MK. Assessment of a respiratory face mask for capturing air pollutants and pathogens including human influenza and rhinoviruses. J Thorac Dis 2018 Mar;10(3):2059-2069. [doi: 10.21037/jtd.2018.03.103]
8. Leung NHL, Chu DKW, Shiu EYC, Chan KH, McDevitt JJ, Hau BJP, et al. Respiratory virus shedding in exhaled breath and efficacy of face masks. Nat Med 2020 May;26(5):676-680 [FREE Full text] [doi: 10.1038/s41591-020-0843-2] [Medline: 32371934]
9. Abeler J, Bäcker M, Buermeyer U, Zillessen H. COVID-19 Contact Tracing and Data Protection Can Go Together. JMIR Mhealth Uhealth 2020 Apr 20;8(4):e19359 [FREE Full text] [doi: 10.2196/19359] [Medline: 32294052]
10. AÕ AIR. URL: https://www.ao-air.com/ [accessed 2022-02-16]
11. LG PuriCare™ Wearable Air Purifier (2nd Gen.). LG. URL: https://www.lg.com/sg/air-purifier/lg-ap551awfa [accessed 2022-04-18]
12. User Resources. ATMOBLUE. URL: https://atmoblue.com/my-atmoblue.html [accessed 2022-02-16]
13. LG PuriCare™ Wearable Air Purifier (2nd Gen.). Amazon. URL: https://www.amazon.com/dp/B08D3JXNB9/ref=emc_b_5_t [accessed 2022-02-16]
14. Philips Fresh Air Mask. Philips. URL: https://www.philips.com/a-w/about/innovation/experience-design/our-work/philips-fresh-air-mask.html [accessed 2022-02-16]
15. Purely Air Purifying Respirator Mask Gray. Xiaomi. URL: https://xiaomi-mi.com/air-and-water-purifiers/xiaomi-purely-air-purifying-respirator-mask-gray/ [accessed 2022-02-16]
16. AIRVISOR Air Cleaning Mask. Gobiz Korea. URL: https://www.gobizkorea.com/user/goods/frontGoodsDetail.do?goods_no=GS2021090606943&goodsNms= [accessed 2022-04-18]
17. CELLRETURN CX9 (Half Mask). CELLRETURN. URL: https://cellreturn.sg/products/cellreturn-cx9-half-mask-white [accessed 2022-02-16]
18. Razer Zephyr Transparent RGB Wearable Air Purifier with Air Filters. Razer. URL: https://www.razer.com/gear-accessories/razer-zephyr/RC81-03870128-R3M1 [accessed 2022-02-16]
19. CLIU. URL: https://cliu.it/ [accessed 2022-04-18]
20. C-Face smart mask. Donut Robotics. URL: https://www.donutrobotics.com/c-mask [accessed 2022-02-16]
21. LED Face Mask, Light Up Mask For Adults, Bluetooth Smartphone App, Rechargeable, Digital Electronic Display, Programmable, For Halloween, Parties and Christmas. Amazon. URL: https://www.amazon.com/TrendyNow365-Customizable-Programmable-Illuminating-Smartphone/dp/B08KT6S76Z [accessed 2022-02-16]
22. Kalavakonda RR, Masna NVR, Mandal S, Bhunia S. A smart mask for active defense against airborne pathogens. Sci Rep 2021 Oct 07;11(1):19910 [FREE Full text] [doi: 10.1038/s41598-021-99150-x] [Medline: 34620887]
23. Nguyen PQ, Soenksen LR, Donghia NM, Angenent-Mari NM, de Puig H, Huang A, et al. Wearable materials with embedded synthetic biology sensors for biomolecule detection. Nat Biotechnol 2021 Nov 28;39(11):1366-1374. [doi: 10.1038/s41587-021-00950-3] [Medline: 34183860]
24. Ye Z, Ling Y, Yang M, Xu Y, Zhu L, Yan Z, et al. A Breathable, Reusable, and Zero-Power Smart Face Mask for Wireless Cough and Mask-Wearing Monitoring. ACS Nano 2022 Mar 17;16(4):5874-5884. [doi: 10.1021/acsnano.1c11041] [Medline: 35298138]
25. Fois A, Tocco F, Dell'Osa A, Melis L, Bertelli U, Concu A. Innovative Smart Face Mask to Protect Workers from COVID-19 Infection. 2021 Presented at: 2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA); June 23-25, 2021; Lausanne. [doi: 10.1109/memea52024.2021.9478739]
26. Lazaro M, Lazaro A, Villarino R, Girbau D. Smart Face Mask with an Integrated Heat Flux Sensor for Fast and Remote People's Healthcare Monitoring. Sensors (Basel) 2021 Nov 10;21(22):7472 [FREE Full text] [doi: 10.3390/s21227472] [Medline: 34833547]
27. Escobedo P, Fernández-Ramos MD, López-Ruiz N, Moyano-Rodríguez O, Martínez-Olmos A, Pérez de Vargas-Sansalvador IM, et al. Smart facemask for wireless CO monitoring. Nat Commun 2022 Jan 10;13(1):72 [FREE Full text] [doi: 10.1038/s41467-021-27733-3] [Medline: 35013232]

28. Zhong J, Li Z, Takakuwa M, Inoue D, Hashizume D, Jiang Z, et al. Smart Face Mask Based on an Ultrathin Pressure Sensor for Wireless Monitoring of Breath Conditions. Adv Mater 2022 Feb 30;34(6):e2107758. [doi: 10.1002/adma.202107758] [Medline: 34706136]
29. Yang L, Wang H, Yuan W, Li Y, Gao P, Tiwari N, et al. Wearable Pressure Sensors Based on MXene/Tissue Papers for Wireless Human Health Monitoring. ACS Appl Mater Interfaces 2021 Dec 22;13(50):60531-60543. [doi: 10.1021/acsami.1c22001] [Medline: 34894673]
30. Kim N, Wei J, Ying J, Zhang H, Moon S, Choi J. A customized smart medical mask for healthcare personnel. 2020 Presented at: 2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM); December 14-17, 2020; Singapore. [doi: 10.1109/ieem45057.2020.9309863]
31. Pan L, Wang C, Jin H, Li J, Yang L, Zheng Y, et al. Lab-on-Mask for Remote Respiratory Monitoring. ACS Mater Lett 2020 Aug 07;2(9):1178-1181 [FREE Full text] [doi: 10.1021/acsmaterialslett.0c00299] [Medline: 34192277]
32. Gravina R, López R, Ordoñez-Ordoñez P. FaceMask: a Smart Personal Protective Equipment for Compliance Assessment of Best Practices to Control Pandemic. 2021 Presented at: 2021 IEEE 2nd International Conference on Human-Machine Systems (ICHMS); September 8-10, 2021; Magdeburg. [doi: 10.1109/ichms53169.2021.9582635]
33. Curtiss A, Rothrock B, Bakar A, Arora N, Huang J, Englhardt Z, et al. FaceBit: Smart Face Masks Platform. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. New York, NY: Association for Computing Machinery; Dec 27, 2021.
34. Fischer H, Wittmann D, Baucells CA, Zhou B, Joost G, Lukowicz P. Masquare: A Functional Smart Mask Design for Health Monitoring. 2021 Presented at: UbiComp '21: The 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing; Virtual; September 21-26, 2021. [doi: 10.1145/3460421.3478831]
35. Fischer JB, Kobayashi Frisk L, Scholkmann F, Delgado-Mederos R, Mayos M, Durduran T. Cerebral and systemic physiological effects of wearing face masks in young adults. Proc Natl Acad Sci U S A 2021 Oct 12;118(41) [FREE Full text] [doi: 10.1073/pnas.2109111118] [Medline: 34607955]
36. Park S, Han J, Yeon YM, Kang NY, Kim E, Suh BF. Long-term effects of face masks on skin characteristics during the COVID-19 pandemic. Skin Res Technol 2022 Jan 19;28(1):153-161 [FREE Full text] [doi: 10.1111/srt.13107] [Medline: 34668244]
37. Elisheva R. Adverse Effects of Prolonged Mask Use among Healthcare Professionals during COVID-19. J Infect Dis Epidemiol 2020 Jun 01;6(3). [doi: 10.23937/2474-3658/1510130]
38. O'Connor Z, Huellewig D, Sithiyopasakul P, Morris J, Gan C, Ballard D. 3D printed mask extenders as a supplement to isolation masks to relieve posterior auricular discomfort: an innovative 3D printing response to the COVID-19 pandemic. 3D Print Med 2020 Sep 29;6(1):27 [FREE Full text] [doi: 10.1186/s41205-020-00080-7] [Medline: 32990872]
39. Edwards NJ, Colder B, Sullivan J, Naramore L. A Practical Approach to Indoor Air Quality for Municipal Public Health and Safety. OJPS 2021;11(01):176-191. [doi: 10.4236/ojps.2021.111012]
40. Myers B, Nahal J, Yang C, Brown L, Ghiasi S, Knoesen A. Towards data-driven pre-operative evaluation of lung cancer patients: the case of smart mask. 2016 Presented at: 2016 IEEE Wireless Health (WH); October 25-27, 2016; Bethesda, MD. [doi: 10.1109/wh.2016.7764569]
41. Saasa V, Malwela T, Beukes M, Mokgotho M, Liu C, Mwakikunga B. Sensing Technologies for Detection of Acetone in Human Breath for Diabetes Diagnosis and Monitoring. Diagnostics (Basel) 2018 Jan 31;8(1):12 [FREE Full text] [doi: 10.3390/diagnostics8010012] [Medline: 29385067]
42. Landini BE, Bravard ST. Breath Acetone Concentration Measured Using a Palm-Size Enzymatic Sensor System. IEEE Sensors J 2009 Dec;9(12):1802-1807. [doi: 10.1109/jsen.2009.2033305]
43. Wang D, Zhang F, Prabhakar A, Qin X, Forzani ES, Tao N. Colorimetric Sensor for Online Accurate Detection of Breath Acetone. ACS Sens 2021 Feb 26;6(2):450-453. [doi: 10.1021/acssensors.0c02025] [Medline: 33210907]
44. Panes-Ruiz LA, Riemenschneider L, Al Chawa MM, Löffler M, Rellinghaus B, Tetzlaff R, et al. Selective and self-validating breath-level detection of hydrogen sulfide in humid air by gold nanoparticle-functionalized nanotube arrays. Nano Res 2022 Sep 02;15(3):2512-2521 [FREE Full text] [doi: 10.1007/s12274-021-3771-7] [Medline: 34493951]
45. Singh E, Meyyappan M, Nalwa HS. Flexible Graphene-Based Wearable Gas and Chemical Sensors. ACS Appl Mater Interfaces 2017 Oct 11;9(40):34544-34586. [doi: 10.1021/acsami.7b07063] [Medline: 28876901]
46. Ahmed AM, Mehaney A, Elsayed HA. Detection of toluene traces in exhaled breath by using a 1D PC as a biomarker for lung cancer diagnosis. Eur. Phys. J. Plus 2021 Jun 04;136(6). [doi: 10.1140/epjp/s13360-021-01621-7]
47. Aasi A, Aghaei SM, Panchapakesan B. A density functional theory study on the interaction of toluene with transition metal decorated carbon nanotubes: a promising platform for early detection of lung cancer from human breath. Nanotechnology 2020 Oct 09;31(41):415707. [doi: 10.1088/1361-6528/ab9da9] [Medline: 32554899]
48. Adiguzel Y, Kulah H. Breath sensors for lung cancer diagnosis. Biosens Bioelectron 2015 Mar 15;65:121-138. [doi: 10.1016/j.bios.2014.10.023] [Medline: 25461148]
49. Mohan PM, Nagarajan V, Das SR. Stress measurement from wearable photoplethysmographic sensor using heart rate variability data. 2016 Presented at: 2016 International Conference on Communication and Signal Processing (ICCSP); April 6-8, 2016; Melmaruvathur. [doi: 10.1109/ICCSP.2016.7754331]

50. Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology. Heart Rate Variability. Circulation 1996 Mar;93(5):1043-1065. [doi: 10.1161/01.cir.93.5.1043]
51. Hjortskov N, Rissén D, Blangsted AK, Fallentin N, Lundberg U, Søgaard K. The effect of mental stress on heart rate variability and blood pressure during computer work. Eur J Appl Physiol 2004 Jun 1;92(1-2):84-89. [doi: 10.1007/s00421-004-1055-z] [Medline: 14991326]
52. Holz C, Wang EJ. Glabella: Continuously Sensing Blood Pressure Behavior using an Unobtrusive Wearable Device. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. New York, NY: Association for Computing Machinery; 2017:1-23.
53. Mena LJ, Félix VG, Ostos R, González AJ, Martínez-Peláez R, Melgarejo JD, et al. Mobile Personal Health Care System for Noninvasive, Pervasive, and Continuous Blood Pressure Monitoring: Development and Usability Study. JMIR Mhealth Uhealth 2020 Jul 20;8(7):e18012 [FREE Full text] [doi: 10.2196/18012] [Medline: 32459642]
54. Shustak S, Inzelberg L, Steinberg S, Rand D, David Pur M, Hillel I, et al. Home monitoring of sleep with a temporary-tattoo EEG, EOG and EMG electrode array: a feasibility study. J Neural Eng 2019 Apr 12;16(2):026024. [doi: 10.1088/1741-2552/aafa05] [Medline: 30566912]
55. Paul GM, Fan C, Torah R, Kai Y, Beeby S, Tudor J. A Smart Textile Based Facial EMG and EOG Computer Interface. IEEE Sensors J 2014 Feb;14(2):393-400. [doi: 10.1109/jsen.2013.2283424]
56. Thompson D, Mackenzie IG, Leuthold H, Filik R. Emotional responses to irony and emoticons in written language: Evidence from EDA and facial EMG. Psychophysiology 2016 Jul 17;53(7):1054-1062. [doi: 10.1111/psyp.12642] [Medline: 26989844]
57. Udovičić G, Đerek J, Russo M, Sikora M. Wearable emotion recognition system based on GSR and PPG signals. 2017 Presented at: MM '17: ACM Multimedia Conference; October 23, 2017; Mountain View, CA. [doi: 10.1145/3132635.3132641]
58. Maffei A, Angrilli A. Spontaneous eye blink rate: An index of dopaminergic component of sustained attention and fatigue. Int J Psychophysiol 2018 Jan;123:58-63. [doi: 10.1016/j.ijpsycho.2017.11.009] [Medline: 29133149]
59. Gjoreski HI, Mavridou II, Fatoorechi M, Kiprijanovska I, Gjoreski M, Cox G, et al. emteqPRO: Face-mounted Mask for Emotion Recognition and Affective Computing. 2021 Presented at: UbiComp '21: The 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 21-26, 2021; Virtual, USA. [doi: 10.1145/3460418.3479276]
60. Yang G, Jiang M, Ouyang W, Ji G, Xie H, Rahmani AM, et al. IoT-Based Remote Pain Monitoring System: From Device to Cloud Platform. IEEE J Biomed Health Inform 2018 Nov;22(6):1711-1719. [doi: 10.1109/jbhi.2017.2776351]
61. Severin IC. Head Posture Monitor Based On 3 IMU Sensors: Consideration Toward Healthcare Application. 2020 Presented at: 2020 International Conference on e-Health and Bioengineering (EHB); October 29-30, 2020; Iasi. [doi: 10.1109/ehb50910.2020.9280106]
62. Pigliautile I, Marseglia G, Pisello AL. Investigation of CO2 Variation and Mapping Through Wearable Sensing Techniques for Measuring Pedestrians' Exposure in Urban Areas. Sustainability 2020 May 11;12(9):3936. [doi: 10.3390/su12093936]
63. Duan Z, Jiang Y, Yan M, Wang S, Yuan Z, Zhao Q, et al. Facile, Flexible, Cost-Saving, and Environment-Friendly Paper-Based Humidity Sensor for Multifunctional Applications. ACS Appl Mater Interfaces 2019 Jun 19;11(24):21840-21849. [doi: 10.1021/acsami.9b05709] [Medline: 31135126]
64. Jia Y, Heiss M, Fu Q, Gay N. A prototype RFID humidity sensor for built environment monitoring. 2008 Presented at: 2008 International Workshop on Education Technology and Training & 2008 International Workshop on Geoscience and Remote Sensing; December 21-22, 2008; Shanghai. [doi: 10.1109/ettandgrs.2008.35]
65. Buchberger A, Peterka S, Coclite A, Bergmann A. Fast Optical Humidity Sensor Based on Hydrogel Thin Film Expansion for Harsh Environment. Sensors (Basel) 2019 Feb 26;19(5):999 [FREE Full text] [doi: 10.3390/s19050999] [Medline: 30813631]
66. Hong S, Kwon H, Choi SH, Park KS. Intelligent system for drowsiness recognition based on ear canal electroencephalography with photoplethysmography and electrocardiography. Inf Sci 2018 Jul;453:302-322. [doi: 10.1016/j.ins.2018.04.003]
67. Yapici MK, Alkhidir TE. Intelligent Medical Garments with Graphene-Functionalized Smart-Cloth ECG Sensors. Sensors (Basel) 2017 Apr 16;17(4) [FREE Full text] [doi: 10.3390/s17040875] [Medline: 28420158]
68. Ikeda S, Ishimura H, Matsumura M. Non-restrictive measurement of pulse transit time using ECG sensor and PPG sensor mounted on the neckband. Trans Jpn Soc Med Biol Eng 2013;51:R-2. [doi: 10.11239/jsmbe.51.R-2]
69. Shaukat N, Ali DM, Razzak J. Physical and mental health impacts of COVID-19 on healthcare workers: a scoping review. Int J Emerg Med 2020 Jul 20;13(1):40 [FREE Full text] [doi: 10.1186/s12245-020-00299-5] [Medline: 32689925]
70. Mishra T, Wang M, Metwally AA, Bogu GK, Brooks AW, Bahmani A, et al. Pre-symptomatic detection of COVID-19 from smartwatch data. Nat Biomed Eng 2020 Dec 18;4(12):1208-1220. [doi: 10.1038/s41551-020-00640-6] [Medline: 33208926]
71. Lane N, Miluzzo E, Lu H, Peebles D, Choudhury T, Campbell A. A survey of mobile phone sensing. IEEE Commun Mag 2010 Sep;48(9):140-150. [doi: 10.1109/mcom.2010.5560598]
72. Khan WZ, Xiang Y, Aalsalem MY, Arshad Q. Mobile Phone Sensing Systems: A Survey. IEEE Commun Surv Tutorials 2013;15(1):402-427. [doi: 10.1109/surv.2012.031412.00077]

73. Sikimić M, Amović M, Vujović V, Suknović B, Manjak D. An Overview of Wireless Technologies for IoT Network. 2020 Presented at: 2020 19th International Symposium INFOTEH-JAHORINA (INFOTEH); March 18-20, 2020; East Sarajevo. [doi: 10.1109/infoteh48170.2020.9066337]
74. Masna N, Kalavakonda R, Dizon R, Bhunia S. Smart and Connected Mask for Protection beyond the Pandemic. 2021 Presented at: 2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS); August 9-11, 2021; Lansing, MI. [doi: 10.1109/mwscas47672.2021.9531802]
75. Hyysalo J, Dasanayake S, Hannu J, Schuss C, Rajanen M, Leppänen T, et al. Smart mask – Wearable IoT solution for improved protection and personal health. Internet of Things 2022 May;18:100511. [doi: 10.1016/j.iot.2022.100511]
76. Mustapää T, Nikander P, Hutzschenreuter D, Viitala R. Metrological Challenges in Collaborative Sensing: Applicability of Digital Calibration Certificates. Sensors (Basel) 2020 Aug 21;20(17):4730 [FREE Full text] [doi: 10.3390/s20174730] [Medline: 32825715]
77. Al‐Turjman F, Baali I. Machine learning for wearable IoT‐based applications: A survey. Trans Emerging Tel Tech 2019 May 16:e3635. [doi: 10.1002/ett.3635]
78. Benis A, Tamburis O, Chronaki C, Moen A. One Digital Health: A Unified Framework for Future Health Ecosystems. J Med Internet Res 2021 Feb 05;23(2):e22189 [FREE Full text] [doi: 10.2196/22189] [Medline: 33492240]
79. Runeson P, Olsson T, Linåker J. Open Data Ecosystems — An empirical investigation into an emerging industry collaboration concept. J Syst Softw 2021 Dec;182:111088. [doi: 10.1016/j.jss.2021.111088]
80. Chen L, Zhao Y, Ye P, Zhang J, Zou J. Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers. Expert Syst Appl 2017 Nov;85:279-291. [doi: 10.1016/j.eswa.2017.01.040]
81. Mihalick VL, Canada JM, Arena R, Abbate A, Kirkman DL. Cardiopulmonary exercise testing during the COVID-19 pandemic. Prog Cardiovasc Dis 2021 Jul;67:35-39. [doi: 10.1016/j.pcad.2021.04.005] [Medline: 33964290]
82. Westmattelmann D, Grotenhermen J, Sprenger M, Schewe G. The show must go on - virtualisation of sport events during the COVID-19 pandemic. Eur J Inf Syst 2020 Dec 06;30(2):119-136. [doi: 10.1080/0960085x.2020.1850186]
83. Chiu SW, Wang JH, Chang KH, Chang TH, Wang CM, Chang CL, et al. A Fully Integrated Nose-on-a-Chip for Rapid Diagnosis of Ventilator-Associated Pneumonia. IEEE Trans Biomed Circuits Syst 2014 Dec;8(6):765-778. [doi: 10.1109/tbcas.2014.2377754]
84. Ingle MA, Talmale GR. Respiratory Mask Selection and Leakage Detection System Based on Canny Edge Detection Operator. Procedia Computer Science 2016;78:323-329. [doi: 10.1016/j.procs.2016.02.064]
85. Kodam S, Bharathgoud N, Ramachandran B. A review on smart wearable devices for soldier safety during battlefield using WSN technology. Materials Today: Proceedings 2020;33:4578-4585. [doi: 10.1016/j.matpr.2020.08.191]
86. Conroy B, Silva I, Mehraei G, Damiano R, Gross B, Salvati E, et al. Real-time infection prediction with wearable physiological monitoring and AI to aid military workforce readiness during COVID-19. Sci Rep 2022 Mar 08;12(1):3797 [FREE Full text] [doi: 10.1038/s41598-022-07764-6] [Medline: 35260671]
87. Sundararajan A, Sarwat AI, Pons A. A Survey on Modality Characteristics, Performance Evaluation Metrics, and Security for Traditional and Wearable Biometric Systems. ACM Comput Surv 2020 Mar 31;52(2):1-36. [doi: 10.1145/3309550]
88. Kamišalić A, Fister I, Turkanović M, Karakatič S. Sensors and Functionalities of Non-Invasive Wrist-Wearable Devices: A Review. Sensors (Basel) 2018 May 25;18(6) [FREE Full text] [doi: 10.3390/s18061714] [Medline: 29799504]
89. Sadeghi Bazargani H, Saadati M, Rezapour R, Abedi L. Determinants and barriers of helmet use in Iranian motorcyclists: a systematic review. J Inj Violence Res 2017 Jan 13;9(1) [FREE Full text] [doi: 10.5249/jivr.v9i1.890] [Medline: 28042961]
90. Faryabi J, Rajabi M, Alirezaee S. Evaluation of the use and reasons for not using a helmet by motorcyclists admitted to the emergency ward of shahid bahonar hospital in kerman. Arch Trauma Res 2014 Sep 23;3(3):e19122. [doi: 10.5812/atr.19122] [Medline: 25599066]
91. Pousette MW, Lo Martire R, Linder J, Kristoffersson M, Äng BO. Neck Muscle Strain in Air Force Pilots Wearing Night Vision Goggles. Aerospace Medicine and Human Performance 2016 Nov 01;87(11):928-932. [doi: 10.3357/amhp.4579.2016]
92. Cloet A, Griffin L, Yu M, Durfee W. Design considerations for protective mask development: A remote mask usability evaluation. Appl Ergon 2022 Jul;102:103751. [doi: 10.1016/j.apergo.2022.103751] [Medline: 35339761]
93. Kumagai J, Kembel L, Ewashko T. Usability Review of Mask Extenders and Ear Savers. In: Proceedings of the 21st Congress of the International Ergonomics Association (IEA 2021). IEA 2021. Lecture Notes in Networks and Systems. Cham: Springer; 2021.
94. Saeidnia HR, Karajizadeh M, Mohammadzadeh Z, Abdoli S, Hassanzadeh M. Usability Evaluation of the Mask Mobile Application: The Official Application of the Iranian Government. Iran J Med Microbiol 2022 Jan 01;16(1):49-55. [doi: 10.30699/ijmm.16.1.49]
95. Martin T, Jovanov E, Raskovic D. Issues in wearable computing for medical monitoring applications: a case study of a wearable ECG monitoring device. 2000 Presented at: Fourth International Symposium on Wearable Computers; October 16-17, 2000; Atlanta, GA. [doi: 10.1109/iswc.2000.888463]
96. Kos M, Kramberger I. A Wearable Device and System for Movement and Biometric Data Acquisition for Sports Applications. IEEE Access 2017:1-1. [doi: 10.1109/access.2017.2675538]

97. Liu Y, Wang H, Zhao W, Zhang M, Qin H, Xie Y. Flexible, Stretchable Sensors for Wearable Health Monitoring: Sensing Mechanisms, Materials, Fabrication Strategies and Features. Sensors (Basel) 2018 Feb 22;18(2) [FREE Full text] [doi: 10.3390/s18020645] [Medline: 29470408]
98. Lee J, Kim D, Ryoo H, Shin B. Sustainable Wearables: Wearable Technology for Enhancing the Quality of Human Life. Sustainability 2016 May 11;8(5):466. [doi: 10.3390/su8050466]
99. Cheng JW, Mitomo H. The underlying factors of the perceived usefulness of using smart wearable devices for disaster applications. Telemat Inform 2017 May;34(2):528-539. [doi: 10.1016/j.tele.2016.09.010]
100. Corcoran P. A privacy framework for the Internet of Things. 2016 Presented at: 2016 IEEE 3rd World Forum on Internet of Things (WF-IoT); December 12-14, 2016; Reston, VA. [doi: 10.1109/wf-iot.2016.7845505]
101. Motti V, Caine K. Users' Privacy Concerns About Wearables. 2015 Presented at: International Conference on Financial Cryptography and Data Security; January 26-30, 2015; San Juan. [doi: 10.1007/978-3-662-48051-9_17]
102. Xu B, Xu LD, Cai H, Xie C, Hu J, Bu F. Ubiquitous Data Accessing Method in IoT-Based Information System for Emergency Medical Services. IEEE Trans Ind Inf 2014 May;10(2):1578-1586. [doi: 10.1109/tii.2014.2306382]
103. Jan MA, Khan F, Khan R, Mastorakis S, Menon VG, Alazab M, et al. A Lightweight Mutual Authentication and Privacy-preservation Scheme for Intelligent Wearable Devices in Industrial-CPS. IEEE Trans Industr Inform 2021 Aug;17(8):5829-5839. [doi: 10.1109/tii.2020.3043802] [Medline: 33981186]
104. Guo C, Tian P, Choo KR. Enabling Privacy-Assured Fog-Based Data Aggregation in E-Healthcare Systems. IEEE Trans Ind Inf 2021 Mar;17(3):1948-1957. [doi: 10.1109/tii.2020.2995228]
105. Nguyen D, Marcu G, Hayes G, Truong K, Scott J, Langheinrich M, et al. Encountering SenseCam: personal recording technologies in everyday life. 2009 Presented at: Ubicomp '09: The 11th International Conference on Ubiquitous Computing; September 30 - October 3, 2009; Orlando, FL. [doi: 10.1145/1620545.1620571]
106. Denning T, Dehlawi Z, Kohno T. In situ with bystanders of augmented reality glasses: Perspectives on recording and privacy-mediating technologies. 2014 Presented at: CHI '14: CHI Conference on Human Factors in Computing Systems; April 26 - May 1, 2014; Toronto, ON. [doi: 10.1145/2556288.2557352]
107. Zhang K, Ni J, Yang K, Liang X, Ren J, Shen XS. Security and Privacy in Smart City Applications: Challenges and Solutions. IEEE Commun Mag 2017 Jan;55(1):122-129. [doi: 10.1109/mcom.2017.1600267cm]
108. Zhang C, Shahriar H, Riad A. Security and Privacy Analysis of Wearable Health Device. 2020 Presented at: 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC); July 13-17, 2020; Madrid. [doi: 10.1109/compsac48688.2020.00044]

Abbreviations

EDA: electrodermal activity
EMG: electromyography
EOG: electrooculography
IoT: Internet of Things
KAIST: Korea Advanced Institute of Science & Technology
ML: machine Learning
NFC: near-field communication
PPG: photoplethysmography

Edited by L Buis; submitted 09.04.22; peer-reviewed by J Hester, KW Kim; comments to author 29.04.22; revised version received 20.05.22; accepted 08.06.22; published 21.06.22.

Please cite as:
Lee P, Kim H, Kim Y, Choi W, Zitouni MS, Khandoker A, Jelinek HF, Hadjileontiadis L, Lee U, Jeong Y
Beyond Pathogen Filtration: Possibility of Smart Masks as Wearable Devices for Personal and Group Health and Safety Management
JMIR Mhealth Uhealth 2022;10(6):e38614
URL: https://mhealth.jmir.org/2022/6/e38614
doi:10.2196/38614
PMID:35679029

Review

Impact of the Moderating Effect of National Culture on Adoption Intention in Wearable Health Care Devices: Meta-analysis

Zhenming Zhang[1], PhD; Enjun Xia[1], PhD; Jieping Huang[1], PhD

School of Management and Economics, Beijing Institute of Technology, Beijing, China

Corresponding Author:
Jieping Huang, PhD
School of Management and Economics
Beijing Institute of Technology
5 South Zhongguancun Street, Haidian District
Beijing, 100081
China
Phone: 86 139 1085 0628
Fax: 86 139 1085 0628
Email: cindy@bit.edu.cn

Abstract

Background: Wearable health care devices have not yet been commercialized on a large scale. Additionally, people in different countries have different utilization rates. Therefore, more in-depth studies on the moderating effect of national culture on adoption intention in wearable health care devices are necessary.

Objective: This study aims to explore the summary results of the relationships between perceived usefulness and perceived ease of use with adoption intention in wearable health care devices and the impact of the moderating effect of national culture on these two relationships.

Methods: We searched for studies published before September 2021 in the Web of Science, EBSCO, Engineering Village, China National Knowledge Infrastructure, IEEE Xplore, and Wiley Online Library databases. CMA (version 2.0, Biostat Inc) software was used to perform the meta-analysis. We conducted publication bias and heterogeneity tests on the data. The random-effects model was used to estimate the main effect size, and a sensitivity analysis was conducted. A meta-regression analysis was used to test the moderating effect of national culture.

Results: This meta-analysis included 20 publications with a total of 6128 participants. Perceived usefulness (r=0.612, P<.001) and perceived ease of use (r=0.462, P<.001) positively affect adoption intention. The relationship between perceived usefulness and adoption intention is positively moderated by individualism/collectivism (β=.003, P<.001), masculinity/femininity (β=.008, P<.001) and indulgence/restraint (β=.005, P<.001), and negatively moderated by uncertainty avoidance (β=-.005, P<.001). The relationship between perceived ease of use and adoption intention is positively moderated by individualism/collectivism (β=.003, P<.001), masculinity/femininity (β=.006, P<.001) and indulgence/restraint (β=.009, P<.001), and negatively moderated by uncertainty avoidance (β=-.004, P<.001).

Conclusions: This meta-analysis provided comprehensive evidence on the positive relationship between perceived usefulness and perceived ease of use with adoption intention and the moderating effect of national culture on these two relationships. Regarding the moderating effect, perceived usefulness and perceived ease of use have a greater impact on adoption intention for people in individualistic, masculine, low uncertainty avoidance, and indulgence cultures, respectively.

(JMIR Mhealth Uhealth 2022;10(6):e30960) doi:10.2196/30960

KEYWORDS

wearable health care devices; national culture; moderating effect; meta-analysis

Introduction

Background

A wearable health care device can be defined as "an autonomous, noninvasive device that can perform specific medical functions such as long-term monitoring or improving health" [1]. The device can detect important vital indicators, such as heart rate, and enables rapid and remote autonomous detection and self-management of arrhythmia. These data can also be transmitted to medical institutions to achieve the purpose of remote health monitoring, thereby effectively reducing the number of patient visits and medical costs [2].

Since the outbreak of COVID-19, people have paid increasing attention to health, and the adoption of wearable health care devices is gradually increasing [3,4], but these devices have not yet been commercialized on a large scale. Therefore, it is necessary to conduct in-depth research on the factors that influence the adoption of wearable health care devices to promote the commercialization of the devices.

Many studies have examined adoption intention toward wearable health care devices [5-7]. These studies have mostly adopted the technology acceptance model (TAM) [8,9] and the unified theory of acceptance and use of technology (UTAUT) [5,6] as the main frameworks. In addition to the variables included in TAM and UTAUT, other variables such as trust [9-11], perceived privacy risk (from the privacy calculus model) [1,12], and consumer innovation (from the theory of innovation diffusion) [9,13] have been considered in the literature. Of the two models, TAM is the most concise and influential model [14] and provides a basis for tracing the influence of external factors on adoption intention. This model discusses the relationship between perceived usefulness, perceived ease of use, and adoption intention [15]. It is easy to understand, with information technology features, a strong theoretical foundation, and sufficient empirical support [16-20].

Studies that used this as the main model to analyze wearable health care device adoption intention, however, did not form a unified understanding, and there were conflicting conclusions on the relationship between perceived ease of use and adoption intention. Many studies have empirically confirmed this relationship [8,21,22]; however, some results have shown that this effect is not obvious [13,23]. Some studies have specifically explored the differences in conclusions caused by moderator variables in population characteristics and focused on the influence of different ages [8,24], genders [9], and experiences [25] on adoption intention in wearable health care devices to promote further commercialization of the devices in people with lower acceptance rates. Moreover, scholars have discovered that national culture also affects wearable health care device adoption intention [6,26], and large differences exist in the utilization rate of wearable health care devices in different countries [27]. Although the study by Meier et al [27] pointed out that under different cultural dimensions there are differences in wearable health care device use, it did not concentrate on how each cultural dimension affects adoption intention.

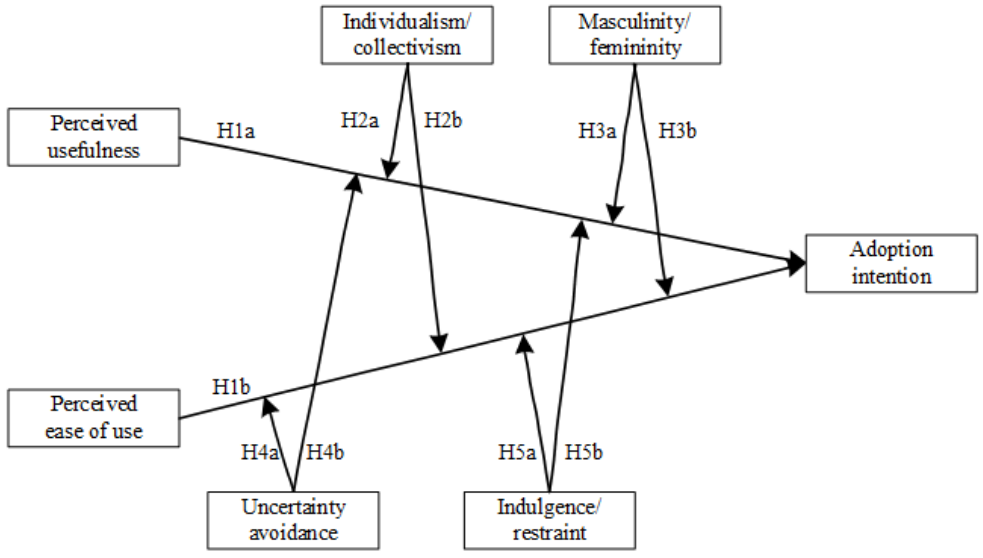
In view of the inconsistent conclusions in the existing studies and the insufficient exploration of the moderating effect of national culture, this study explores summary results of the relationships between perceived usefulness and perceived ease of use in wearable health care device adoption intention and the impact of the moderating effect of national culture on adoption intention by using the meta-analysis method. The results of this study could have implications for global wearable health care device providers in developing and marketing their devices successfully across borders, for effective enhancement of people's health conditions, and for national health agencies to decrease medical expenses.

Theoretical Framework and Hypotheses

Research Framework

The research framework used in this study is presented in Figure 1. We chose TAM as the main model and Hofstede's cultural value dimensions to represent national culture.

Figure 1. Research framework. H: hypothesis.



As mentioned above, TAM is the most concise and influential of the models with a strong theoretical foundation and sufficient empirical support [14,16-20]. The dimensions used to analyze cultural value mainly come from Rokeach [28], Hanson [29], and Hofstede [30]. The dimensions developed by Hofstede are the most recognized and commonly used framework for studying cross-cultural issues on technology adoption [31-34]. The formation process of the value of the cultural dimension has "a rigorous research design, a systematic data collection, and a coherent theory to explain national variations" [35], achieving the aggregation of the properties of individuals as observed within a country. Therefore, every cultural dimension can be treated as a country-level variable [36]. Hofstede's cultural value contains 6 dimensions: power distance, individualism/collectivism, masculinity/femininity, uncertainty avoidance, long-term/short-term orientation, and indulgence/restraint [30]. This study focuses on the moderating effects of 4 of these: individualism/collectivism, masculinity/femininity, uncertainty avoidance, and indulgence/restraint.

First, power distance refers to the degree to which people accept an unequal distribution of power [37]. When commodities can represent the differences in the identity and power of consumers, their purchasing behavior is more susceptible to the influence of power distance [30]. Therefore, power distance is more closely related to luxury purchases in studies on consumer behavior [38,39]. However, a wearable health care device is a health-related and life-oriented product that is not conspicuous. Therefore, power distance has a weak correlation with adoption intention toward wearable health care devices. This paper will not discuss the moderating effect of power distance on the relationships between perceived usefulness and perceived ease of use in adoption intention.

Second, people in a short-term orientation culture value technologies that bring usefulness to current life and work, while people in a long-term orientation culture value technologies that bring usefulness to future life [40]. Wearable health care devices are used not only by patients with chronic diseases [41,42] but also by healthy users for disease prevention [43]. Thus, the importance placed by people in both cultures on perceived usefulness depends on whether the concept is future-oriented or present-oriented. However, the measurement of this concept in the existing literature does not distinguish between these orientations [8,44]; thus, it is difficult to judge the moderating effect of long-term versus short-term orientations on the relationship between perceived usefulness and adoption intention. Moreover, since perceived ease of use is closely related to perceived usefulness [45], the moderating effect of long-term versus short-term orientation on the relationship between perceived ease of use and adoption intention also becomes difficult to judge. Therefore, this study does not analyze and test the moderating effects of long-term and short-term orientation.

Relationships Between Perceived Usefulness and Perceived Ease of Use in Adoption Intention

TAM illustrates the relationships between perceived usefulness and perceived ease of use in adoption intention [46]. Perceived usefulness refers to the degree to which people feel that using technology is helpful to their work and life [15]. Perceived ease of use refers to how much effort people need to use technologies [15]. The relationships between these variables and adoption intention have been proven in many studies related to technology adoption. For example, Hung et al [47] and Wu [48] showed that perceived usefulness and perceived ease of use positively affect the intention to adopt mobile commerce. In our research context, perceived usefulness is not only generally embodied in the improvement of work and life efficiency [13], it is specifically embodied in the improvement of the users' health level [9,44]. These relationships regarding wearable health care devices have been confirmed in multiple studies [21,22]. Thus, we hypothesized the following:

Hypothesis 1a (H1a) and hypothesis 1b (H1b): perceived usefulness and perceived ease of use positively affect adoption intention toward wearable health care devices.

Moderating Effects of Individualism Versus Collectivism

Individualism versus collectivism reflects the degree to which people prefer to care for themselves and their families [30,37,49]. People in an individualistic culture put more emphasis on themselves, while people in a collectivist culture put more emphasis on their families [30,50]. Therefore, people in an individualistic culture value freedom and self-responsibility more and thus value their own health more [30]. This concern for health leads people in individualistic cultures to pay more attention to perceived usefulness of devices before purchase.

People in an individualistic culture are more accustomed to using emerging technologies such as email, online banking, and e-shopping in their daily lives. People from collectivist countries emphasize time spent with family and friends over time spent on the internet [30]. Therefore, people in an individualistic culture might have a higher frequency of using wearable health care devices. If the products are not easy to use, their experiences will be deeply affected. In addition, perceived ease of use positively affects the perceived usefulness of wearable health care devices [45] since perceived ease of use could help realize the function of the devices [51,52]. Moreover, people in an individualistic culture emphasize perceived usefulness more than people in a collectivist culture. Thus, people in an individualistic culture value perceived ease of use more, and we hypothesized the following:

Hypothesis 2a (H2a) and hypothesis 2b (H2b): The higher the degree of individualism, the higher the value placed on perceived usefulness (H2a) and perceived ease of use (H2b) toward adoption intention of wearable health care devices.

Moderating Effects of Masculinity Versus Femininity

Masculinity represents a preference for achievement, heroism, decisiveness, and material rewards for success, while femininity represents cooperation, humility, and quality of life [30]. The perceived usefulness of TAM emphasizes performance improvement and achievement, which is consistent with masculinity [53]. The meaning of achievement changes with time and context. In traditional societies, men pay attention to hunting and fighting, and in modern societies, men value economic achievement [30]. Regarding adoption intention for

wearable health care devices, many people use them to measure sports achievements and enjoy competing with their peers [54]. Therefore, individuals in masculine cultures use wearable health care devices to satisfy their achievement motivation, and they value the perceived usefulness of the wearable health care device more.

People in masculine cultures hope to have challenging jobs to prove their competence and feel a sense of accomplishment, while people in feminine cultures hope to have a safer and higher quality life [30,37]. However, liking challenges does not mean that people in masculine cultures do not value perceived ease of use of wearable health devices. The greatest sense of accomplishment users get from wearable health care devices comes from recording their own sports achievements and competing with others [54] rather than showing they are good at using devices that are not easy to use. The increase in perceived ease of use contributes to the realization of functions of the device, such as functions of measurement, recording, and querying [45,51,52], which can effectively enhance the user's sense of accomplishment. Because people in a masculine culture pay more attention to a sense of accomplishment than people in a feminine culture [30,37], people in a masculine culture also value perceived ease of use more, and we hypothesized the following:

Hypothesis 3a (H3a) and hypothesis 3b (H3b): The higher the degree of masculinity, the higher the value placed on perceived usefulness (H3a) and perceived ease of use (H3b) toward adoption intention of wearable health care devices.

Moderating Effects of Uncertainty Avoidance

People in a culture of high uncertainty avoidance value risk aversion more than people in a culture of low uncertainty avoidance [30]. The adoption of new technologies will bring about new risks, such as privacy risks [1] and imperfect technology [55,56]. This might make people in a high uncertainty avoidance culture hesitate to adopt new technologies. However, wearable health care devices can collect physical health data to control health risks, thereby making health conditions clearer and predictable [57], which is very attractive to people in a culture of high uncertainty avoidance. However, this does not mean that people in a high uncertainty avoidance culture will decide whether to adopt a wearable health care device based on its perceived usefulness. To reduce uncertainty, they are often prepared to engage in risky behavior [49] and are more impulsive [30]. For example, the higher the degree of uncertainty avoidance, the higher the maximum speed limit of a country (region) [30]. In addition, people in a high uncertainty avoidance culture have more concerns about health than people in a culture of low uncertainty avoidance [30]. Therefore, when faced with health-related decisions, people in a culture of high uncertainty avoidance are more likely to ignore meticulous thinking about the perceived usefulness of wearable health care devices and purchase products on impulse.

Regardless of whether people in a culture of high uncertainty avoidance consider the perceived usefulness when purchasing wearable health care devices, their purchase stems from health-related safety requirements [58]. Their need for safety takes precedence over other needs [30], such as the need for comfort and convenience represented by perceived ease of use. Therefore, people in a culture of high uncertainty avoidance pay less attention to the perceived ease of use of wearable health care devices than people in a culture of less uncertainty avoidance. Moreover, because perceived ease of use can improve the perceived usefulness of wearable health care devices [45,51,52] and people in a culture of low uncertainty avoidance are more concerned with perceived usefulness, people in a culture of low uncertainty avoidance perceived ease of use more, and we hypothesized the following:

Hypothesis 4a (H4a) and hypothesis 4b (H4b): The higher the degree of uncertainty avoidance, the less the value placed on perceived usefulness (H4a) and perceived ease of use (H4b) toward adoption intention of wearable health care devices.

Moderating Effects of Indulgence Versus Restraint

People in a culture of indulgence believe that enjoying life and entertainment are basic human needs, and natural desires should be satisfied [30]. People in a culture of restraint believe that human behavior should be restricted by social norms and prohibitions, and enjoying leisure activities, overconsumption, and similar indulgence behaviors are wrong [59]. Therefore, people in a high-indulgence culture are more likely to buy wearable health care devices because of the nonpractical functions of the products such as gamification [60] and innovation [61] rather than practical functions. A larger proportion of people in cultures with greater indulgence claim that their personal health is very good [49]. When people are more confident with their health conditions, they are less likely than people in cultures of restraint to consider perceived usefulness when deciding to purchase health products. Therefore, the greater the indulgence, the lower the value placed on perceived usefulness toward adoption intention of wearable health care devices.

Although people in a restraint culture value perceived usefulness more, and perceived ease of use determines the functional realization of wearable health care devices [45], people in an indulgence culture place more emphasis on perceived ease of use. This may be because people in an indulgence culture prefer pursuing the enjoyment of life [30] over spending time learning to use wearable health care devices. If a device is not easy to use, people in indulgence cultures are less likely to make the purchases. Conversely, people in a restraint culture are taught to be frugal and to limit their desires [30,37], and they believe the pursuit of pleasure is wrong [59]. Therefore, if the perceived usefulness of a device meets their requirements, they will buy and use a device regardless of perceived ease of use, and we hypothesized the following:

Hypothesis 5a (H5a) and hypothesis 5b (H5b): The greater the indulgence, the lower the value placed on perceived usefulness (H5a) and the higher the value placed on perceived ease of use (H5b) toward adoption intention of wearable health care devices.

Methods

Method Selection

Meta-analysis is a quantitative technique that generates a summary effect size for each relationship path [62]. This method

has two functions. First, it helps scholars obtain a summary view of the results [63]. Second, this method is useful for hypothesis testing and moderator analysis [64]. This study used meta-analysis to explore the summary view of the relationships between perceived usefulness and perceived ease of use in adoption intention of wearable health care devices and the impact of the moderating effect of national culture on adoption intention. Therefore, the meta-analysis method is appropriate for this study.

Data Sources and Search Strategy

We conducted a literature search by using keywords such as "wearable*," "health*," "fitness," "wellness," "medical," "accept*," "adopt*," and "intention" to search for studies in the Web of Science, EBSCO, Engineering Village, China National Knowledge Infrastructure, IEEE Xplore, and Wiley Online Library databases. We then manually searched the references of the papers found for additional relevant titles to reduce the influence of publication bias.

Selection Criteria

The study selection criteria were formulated considering the recommendations of Cooper [62] and the aim of this research. Studies included were empirical; reported sample size, correlation coefficient, and country of origin of the surveyed population; were related to adoption intention for wearable health care devices; and surveyed ordinary users and not nursing staff. Studies that did not use TAM or UTAUT as the main model, studies using continuance intention as the dependent variable (because the purpose of this paper is to promote the commercialization of devices rather than the maintenance of users after adoption), multiple studies using the same data (one of the studies would be retained in the paper), and review literature were excluded.

This article treats performance expectation, which belongs to UTAUT, as equivalent to the concept of perceived usefulness, which belongs to TAM. This article treats effort expectation, which belongs to UTAUT, as equivalent to the concept of perceived ease of use, which belongs to TAM. On one hand, other studies have regarded perceived usefulness and performance expectation [65-69] and perceived ease of use and effort expectation [69] as the same concept. On the other, the same results of multiple operations indicate that these operations focus on the same components and can enhance our confidence in the conclusions [62].

Data Extraction

The extracted information included the first author's name, year of publication, sample size, correlation coefficient matrix, and the location of the questionnaire collection. If the author did not report the location, we used the country (region) the authors came from. We got Hofstede's cultural values by searching for that country (region) on the website of Hofstede's cultural dimensions [70]. The required data were extracted independently by two researchers.

Analysis Procedure

The meta-analysis consisted of 4 parts conducted using CMA (version 2.0, Biostat Inc) software. Funnel plots, Egger regression, and Rosenthal fail-safe N tests were used to determine whether publication bias existed [71,72]. The heterogeneity of various items was assessed using a Cochran Q test. When $P<.05$, the heterogeneity test was passed. We also calculated the I^2 statistic, an indicator of heterogeneity in percentages [73].

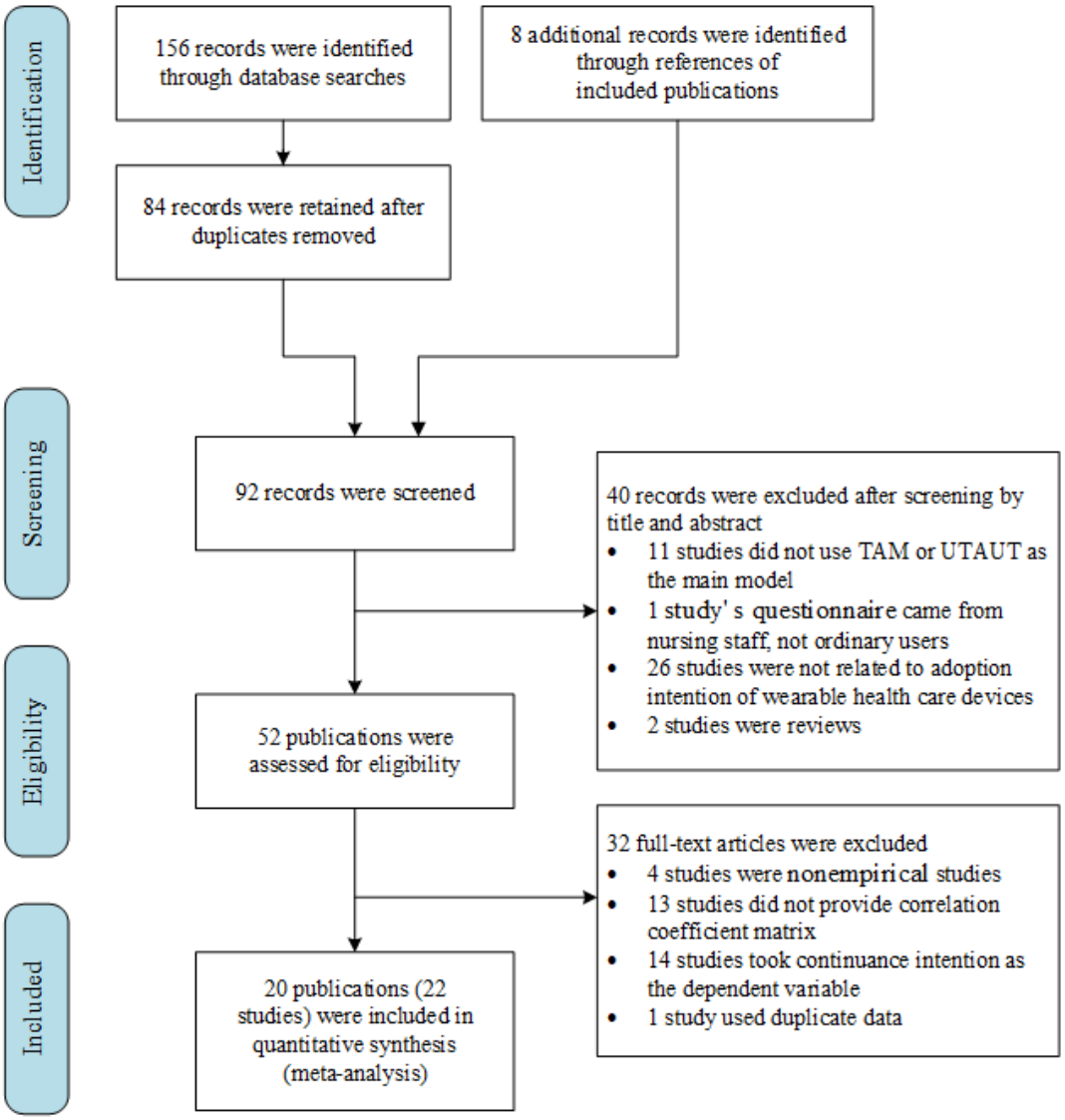
Fixed-effects and random-effects models are the two main methods for calculating effect size [74]. We used the results of the heterogeneity test to select the appropriate model [73]. Because factors such as gender and age might affect the relationships between perceived usefulness and perceived ease of use in adoption intention [46], we used a random-effects model to calculate the main effect size. Sensitivity analysis was conducted to determine whether the elimination of any data item would influence the overall results. We conducted meta-regression analyses to estimate the moderating effects of national culture. For each regression, the correlation coefficient was the dependent variable and the value of the national culture dimension was the independent variable.

Results

Study Selection

A total of 156 papers were found in our search on September 4, 2021. After deduplication, 84 remained, with 8 additional papers identified in the references. Next, 40 papers were excluded based on the titles and abstracts. After reading the full texts of the remaining 52 papers, we deleted 32 that did not meet the selection criteria, with a final total of 20 publications reporting on 22 effect sizes. Two of the 20 papers contained 2 studies. Therefore, 22 studies were included. Figure 2 shows the study flowchart with details.

Figure 2. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart.



Study Characteristics Description

This meta-analysis included 20 publications [5-10,12,13,21,25,26,45,75-82] with 6128 participants. The 20 publications were conducted in 7 countries (regions) and published between 2015 and 2021. The sample size ranged from 100 [5] to 877 [13]. A total of 22 studies analyzed the relationship between perceived usefulness and adoption intention [5-10,12,13,21,25,26,45,75-82], and 18 studies analyzed the relationship between perceived ease of use and adoption intention [5-8,10,12,13,21,25,26,45,75,77,78,80,81,82], and 2 of the studies were from the same publication [25]. The characteristics of the included studies are presented in Multimedia Appendix 1.

Meta-analysis

Publication Bias Test

The results of publication bias test are shown in Table 1, Figure 3, and Figure 4. According to the funnel plot, the studies on the perceived usefulness–adoption intention and perceived ease of use–adoption intention relationships were distributed on either side of the center lines, which indicates that the studies about these relationships do not have publication bias. If the Rosenthal fail-safe N is greater than 5M+10 (M is the number of research papers), publication bias does not exist. Table 1 shows that neither relationship had publication bias. According to the results of the Egger regression intercept, neither relationship had publication bias. Since no publication bias was found using 3 different tests, the main effect sizes of the meta-analysis are considered valid.

Table 1. Results of publication bias test.

Relationship	Rosenthal N	Egger regression intercept				
		Intercept	SE	LL[a]	UL[b]	*P* value
PU[c]-AI[d]	4967	7.489	3.784	–0.405	15.384	.06
PEOU[e]-AI	5047	5.973	4.116	–2.754	14.699	.17

[a]LL: lower limit.

[b]UP: upper limit.

[c]PU: perceived usefulness.

[d]AI: adoption intention.

[e]PEOU: perceived ease of use.

Figure 3. Funnel plot of studies on the perceived usefulness–adoption intention relationship.

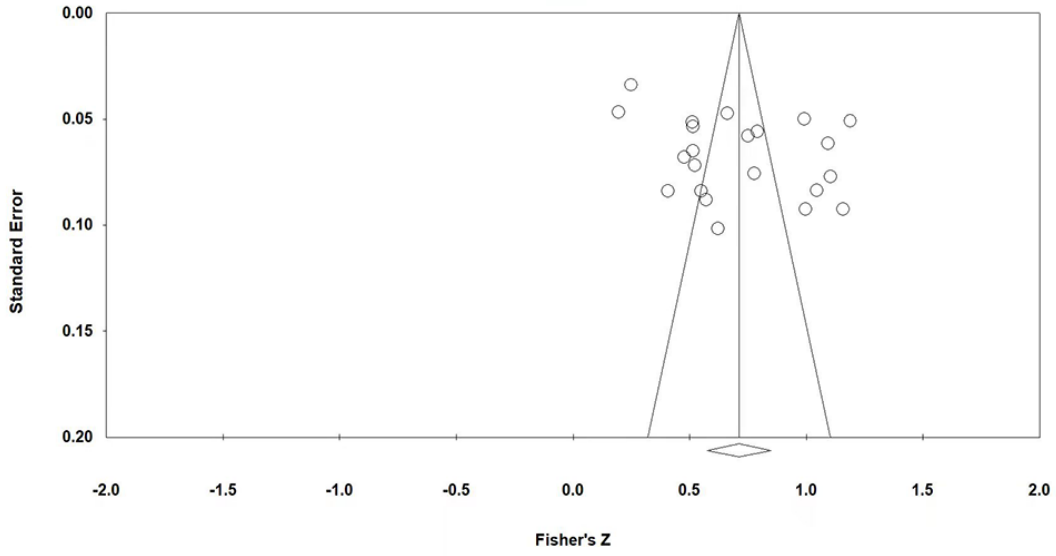
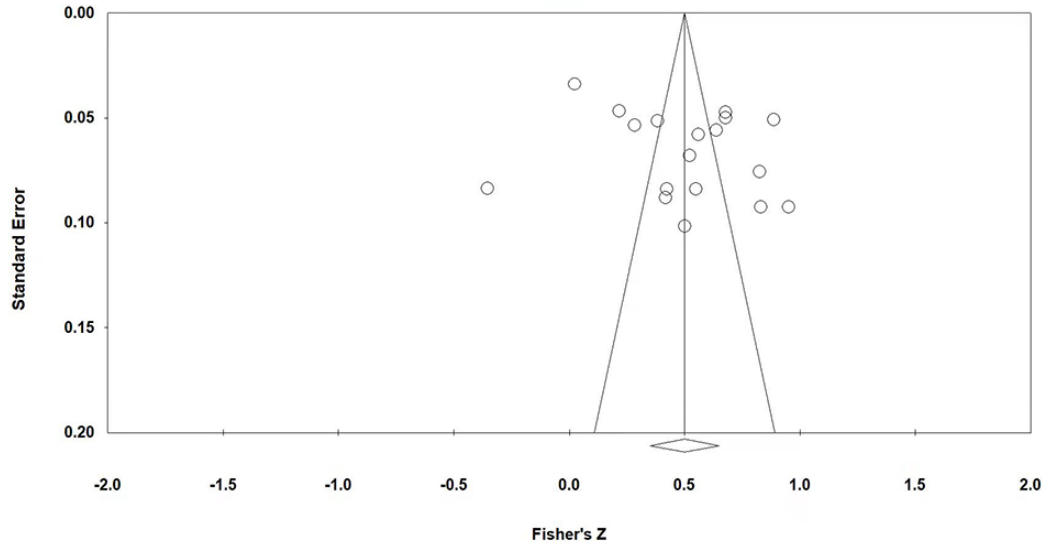


Figure 4. Funnel plot of studies on the perceived ease of use–adoption intention relationship.



Heterogeneity Tests

Table 2 shows that the effect sizes of these studies are heterogeneous. Therefore, it is necessary to test the moderating effect. In addition, the random-effects model should be used when estimating the main effect size.

Table 2. Heterogeneity test results.

Relationship	Heterogeneity			
	Q	*df* (Q)	*P* value	I^2
PU[a]-AI[b]	598.249	21	<.001	96.490
PEOU[c]-AI	495.531	17	<.001	96.569

[a]PU: perceived usefulness.

[b]AI: adoption intention.

[c]PEOU: perceived ease of use.

Estimation of Main Effect Size

The random-effects model was used to test the perceived usefulness–adoption intention and perceived ease of use–adoption intention relationships. Table 3 shows that the perceived usefulness–adoption intention (r=0.612, P<.001) and perceived ease of use–adoption intention (r=0.462, P<.001) relationships were significant. The correlation coefficients are both around 0.5, which means that the perceived usefulness–adoption intention and perceived ease of use–adoption intention relationships have moderately positive correlations [83]. In addition, the results of sensitivity analysis, presented in Figures 5 and 6, showed that the 2 correlation coefficients after any study removed fluctuates between 0.597 and 0.627 (perceived usefulness–adoption intention) and between 0.441 and 0.499 (perceived ease of use–adoption intention), indicating that the results of the meta-analysis have high stability. Therefore, these results confirm hypotheses H1a and H1b.

Table 3. Main effect size estimates.

Hypothesis	Relationship	*k*	Main effect size estimates					Supported
			Point estimate	95% CI		*Z*-value	*P* value	
				LL[a]	UL[b]			
H1a	PU[c]-AI[d]	22	0.612	0.519	0.690	10.224	<.001	Yes
H1b	PEOU[e]-AI	18	0.462	0.336	0.571	6.544	<.001	Yes

[a]LL: lower limit.

[b]UL: upper limit.

[c]PU: perceived usefulness.

[d]AI: adoption intention.

[e]PEOU: perceived ease of use.

Figure 5. Sensitivity analysis results regarding the effect size of the perceived usefulness–adoption intention relationship.

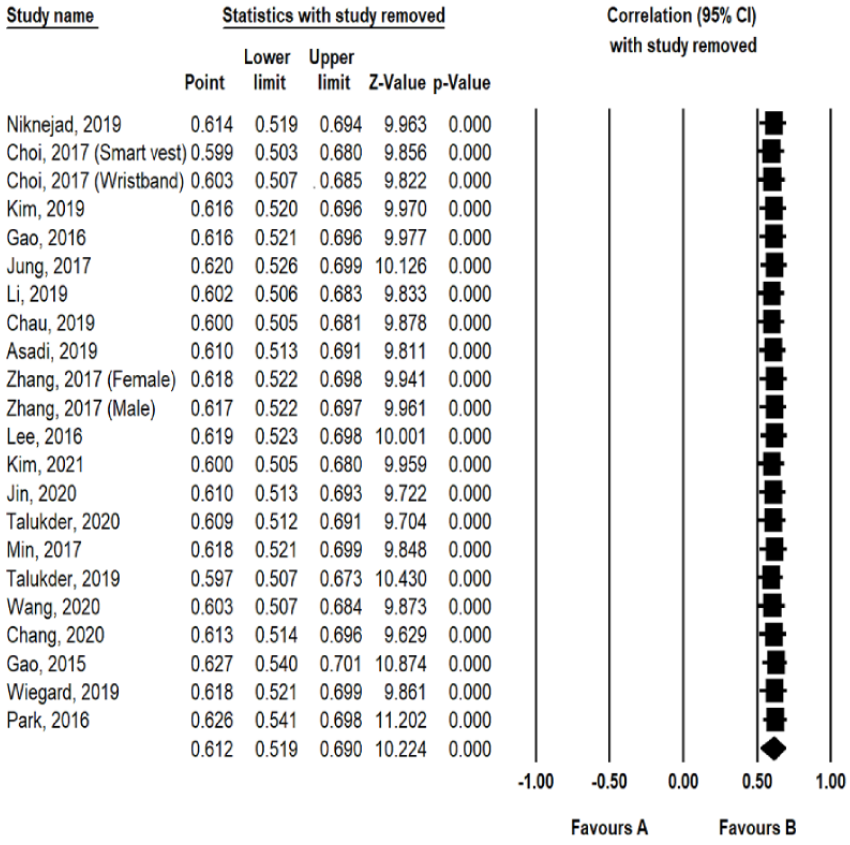
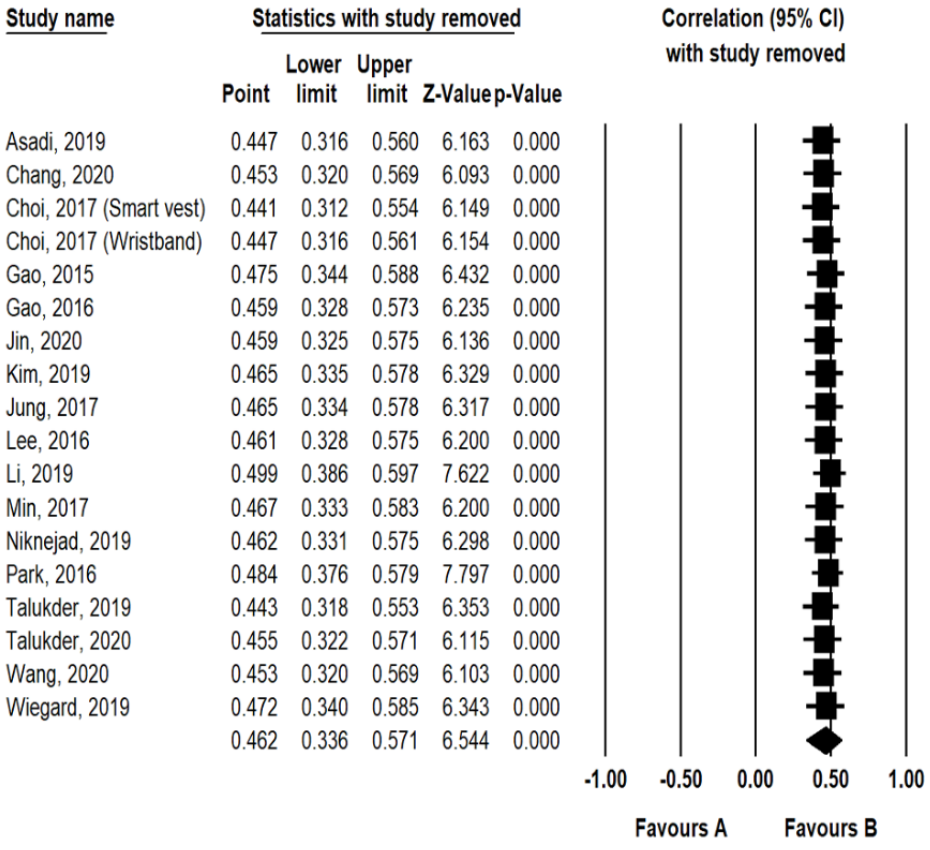


Figure 6. Sensitivity analysis results regarding the effect size of the perceived ease of use–adoption intention relationship.



Estimation of Moderating Effects of National Culture

The results are shown in Table 4. Individualism positively moderates the perceived usefulness–adoption intention (β=.003, *P*<.001) and the perceived ease of use–adoption intention (β=.003, *P*<.001) relationships. These results confirm hypothesis H2a and H2b. Masculinity positively moderates the perceived usefulness–adoption intention (β=.008, *P*<.001) and perceived ease of use–adoption intention (β=.006, *P*<.001) relationships. These results confirm hypotheses H3a and H3b. Uncertainty avoidance negatively moderates the perceived usefulness–adoption intention (β=–.005, *P*<.001) and perceived ease of use–adoption intention (β=–.004, *P*<.001) relationships. These results confirm hypotheses H4a and H4b. Indulgence positively moderates the perceived usefulness–adoption intention (β=.005, *P*<.001) and perceived ease of use–adoption intention (β=.009, *P*<.001) relationships. These results confirm hypothesis H5b but not hypothesis H5a.

The results are summarized in Figure 7. The confirmed hypotheses are represented by a solid line, and the unproven hypotheses are represented by a dashed line.

Table 4. Results of moderating effects of national culture.

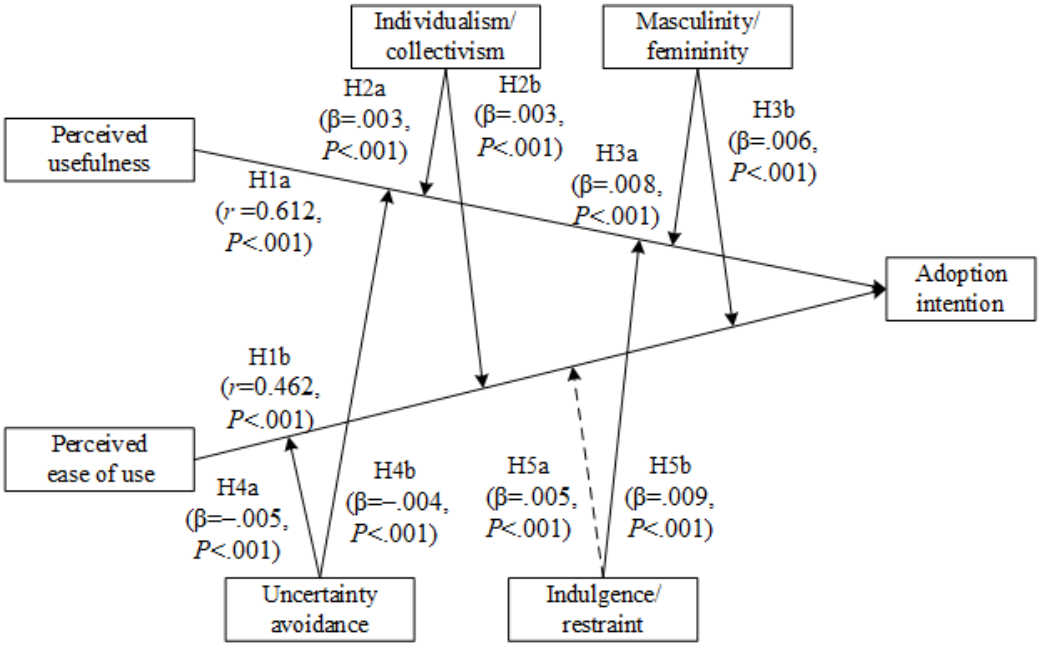
Hypothesis	Relationship	Point estimate	SE	Lower limit	Upper limit	*Z*-value	*P* value	Supported
Individualism/collectivism								
H3a	PU[a]-AI[b]	0.003	0.001	0.002	0.005	4.331	<.001	Yes
H2b	PEOU[c]-AI	0.003	0.001	0.002	0.005	4.095	<.001	Yes
Masculinity/femininity								
H3a	PU-AI	0.008	0.001	0.006	0.01	7.171	<.001	Yes
H3c	PEOU-AI	0.006	0.001	0.004	0.008	5.588	<.001	Yes
Uncertainty avoidance								
H4a	PU-AI	–0.005	0.001	–0.006	–0.004	–9.075	<.001	Yes
H4b	PEOU-AI	–0.004	0.001	–0.005	–0.003	–7.721	<.001	Yes
Indulgence/restraint								
H5a	PU-AI	0.005	0.001	0.003	0.007	5.124	<.001	No
H5b	PEOU-AI	0.009	0.001	0.007	0.011	7.960	<.001	Yes

[a]PU: perceived usefulness.

[b]AI: adoption intention.

[c]PEOU: perceived ease of use.

Figure 7. Meta-analysis results. H: hypothesis.



Discussion

Findings on Main Effects

The results of this study showed that perceived usefulness (H1a) and perceived ease of use (H1b) positively affect adoption intention. These results are consistent with most of the literature on adoption intention in wearable health care devices [9,21]. The results are also consistent with the meta-analysis results in many other research contexts, such as mobile health service adoption [14] and mobile payment adoption [84]. Therefore, the relationships between perceived usefulness and perceived ease of use with adoption intention have once again proved to be robust. Moreover, the results for H1b can help clarify the debate on the relevance direction. This result does not support the uncorrelated result of the relationship between perceived ease of use and adoption intention [63]; thus, the relationship between these two variables should not be ignored in actual work.

Findings on Moderating Effects of National Culture

Gender, age, voluntariness of use, and experience are important moderating variables in UTAUT [46], and gender and age are important in TAM3 [85]. The results of the moderating effects in this paper show that national culture also needs to be a focus in the research context of technology adoption, especially in the context of adoption intention in wearable health care devices. The specific conclusions are as follows:

The results on the moderating effect of individualism/collectivism found that individualism positively moderated the relationship between perceived usefulness and adoption intention (H2a) and the relationship between perceived ease of use and adoption intention (H2b). The test results of H2a and H2b are consistent with the results of Hung and Chou [31] and Zhang et al [86]. H2a states that people in individualistic cultures value personal health more [30], and thus the higher the degree of individualism, the higher the value placed on perceived usefulness toward adoption intention of wearable health care devices (H2a). However, this assumption ignores the fact that an important advantage of wearable health care devices is the implementation of health monitoring and reduction of health risks and costs [2]. People in a collectivist culture are willing to invest less income to maintain health compared to people in an individualistic culture [87]. From this point of view, people in a collectivist culture need devices to protect their health and reduce medical costs. The test result of H2a showed that the importance of mentioned facts in H2a is greater than that of ignored facts. Therefore, H2a is reasonable.

The results on the moderating effect of masculinity/femininity showed that masculinity/femininity positively moderates the influence of perceived usefulness (H3a) and perceived ease of use (H3b) on adoption intention. The test result of H3a is consistent with the findings of Hung and Chou [31], and both results are consistent with the findings of Zhang et al [86]. In our study, people in highly masculine cultures regard health achievements as an aspect of competition. This might be because health is a symbol of strength, which is consistent with the most essential masculine temperament [30]. The test result of H3b is contrary to the findings of Hung and Chou [31]. This result is possible as the perceived ease of use of technologies determines the user experience, and people in a feminine culture value the quality of life more [30]; therefore, people in this culture might value perceived ease of use more. However, when the impact of perceived usefulness on adoption intention is large enough, users who value perceived usefulness will also value perceived ease of use because the perceived ease of use of wearable health care devices could help realize the function of the devices [51,52]. Therefore, the test results of H3b are reasonable.

The results on the moderating effect of uncertainty avoidance showed that uncertainty avoidance negatively moderates the relationship between perceived usefulness (H4a) and perceived ease of use (H4b) with adoption intention. These results are consistent with those of Hung and Chou [31]. The test results for H4a are consistent with the findings of Yoon [88] and Lin [33]; neither study tested H4b. These results show that people in a culture of high uncertainty avoidance are indeed more likely to adopt technologies on impulse and then ignore the perceived usefulness and perceived ease of use of technologies. The negative moderating effect of uncertainty avoidance is easier to understand in this study since health is indeed an important thing for people in a high uncertainty avoidance culture [30] and might lead to irrational buying behaviors.

The results on the moderating effect of indulgence/restraint showed that indulgence strengthens the relationship between perceived ease of use and adoption intention (H5b); however, it does not weaken but strengthens the relationship between perceived usefulness and adoption intention (H5a). H5a states that people in indulgence cultures are less likely to value the perceived usefulness of wearable health care devices because people in such cultures are more likely to consider themselves healthy [30]. However, this reasoning process ignores the fact that people in an indulgence culture consume more junk food and are more obese [30]. In this regard, people in this culture need more wearable health care devices to monitor their health and encourage them to exercise. Thus, indulgence has a positive moderating effect. The test result of H5a showed that people in indulgence cultures rely more on the reality of their health condition when making decisions on adoption intention of wearable health care devices.

Limitations

Our study has several limitations. First, this study focused only on the moderating effect of national culture on the relationship between the variables in TAM and adoption intention. However, the existing literature shows that trust [9-11], perceived privacy risk [1,12], customer innovation [9,13], and other variables affect people's acceptance of wearable health care devices. Subsequent research should further explore the impact of national culture on the relationship between these variables and adoption intention. Second, this study does not discuss the moderating effect of national culture in different subgroups such as gender and age, classic moderating variables in TAM and UTAUT [46,85], because we were unable to obtain more detailed national cultural values of different genders and ages from the official website of Hofstede's cultural dimensions [70]. However, these studies were necessary. For example,

individualism is related to the income levels of individuals [30]. Therefore, the individualism scores of people of different ages in different countries might change when the world's economic structure changes. Thus, it is necessary to conduct subgroup analysis of different ages.

Implications for Practice

The results of this study could have implications for global wearable health care device providers and national health agencies. These results could help wearable health care device providers increase the adoption of the devices worldwide in two ways: guiding providers to develop more attractive and innovative devices by considering cultural factors and steering people toward wearable health care devices at the product sales stage. National health agencies can use these results to persuade people to use the devices for health management, conduct preventive treatment, and decrease medical expenses in the long term.

The application of these conclusions needs to target different national cultures. For example, for people in high masculinity cultures, such as Slovakia, Japan, and Hungary, health care device providers and national health agencies should pay more attention to perceived usefulness in the process of promoting the commercialization of wearable health devices.

When applying these conclusions, we should pay attention to not only the conclusions about the moderating effect of national culture but also the reasons for these conclusions. This can improve the efficiency of the persuasion process. For example, health care device providers and national health agencies should promote user adoption intention by emphasizing the perceived usefulness of the devices for potential users in a high masculinity culture and remind these potential users that they can compare their sport achievements with their peers for motivation using the devices.

Conclusions

This meta-analysis provided comprehensive evidence for the positive relationships between perceived usefulness and perceived ease of use with adoption intention and the moderating effect of national culture on these relationships. Regarding the moderating effect, perceived usefulness and perceived ease of use have a greater impact on adoption intention for people in individualistic, masculine, low uncertainty avoidance and indulgence cultures, respectively.

Acknowledgments

This research was funded by grant 7217020 from the National Natural Science Foundation of China.

Authors' Contributions

ZZ developed the original idea for this study, wrote the original draft, and revised the manuscript. EX performed the data analysis and data presentation. JH developed the original idea and supervised the research project. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of the included studies.
[DOCX File , 23 KB - mhealth_v10i6e30960_app1.docx]

References

1. Li H, Wu J, Gao Y, Shi Y. Examining individuals' adoption of healthcare wearable devices: an empirical study from privacy calculus perspective. Int J Med Inform 2016 Apr;88:8-17. [doi: 10.1016/j.ijmedinf.2015.12.010] [Medline: 26878757]
2. Lee SY, Lee K. Factors that influence an individual's intention to adopt a wearable healthcare device: the case of a wearable fitness tracker. Technol Forecast Soc Change 2018 Apr;129:154-163 [FREE Full text] [doi: 10.1016/j.techfore.2018.01.002]
3. Choudhury A, Asan O. Impact of using wearable devices on psychological Distress: analysis of the health information national Trends survey. Int J Med Inform 2021 Dec;156:104612. [doi: 10.1016/j.ijmedinf.2021.104612] [Medline: 34649113]
4. Ueafuea K, Boonnag C, Sudhawiyangkul T, Leelaarporn P, Gulistan A, Chen W, et al. Potential applications of mobile and wearable devices for psychological support during the COVID-19 pandemic: a review. IEEE Sensors J 2021 Mar 15;21(6):7162-7178. [doi: 10.1109/jsen.2020.3046259]
5. Niknejad N, Hussin ARC, Ghani I, Ganjouei FA. A confirmatory factor analysis of the behavioral intention to use smart wellness wearables in Malaysia. Univ Access Inf Soc 2019 Jul 2;19(3):633-653. [doi: 10.1007/s10209-019-00663-0]
6. Talukder MS, Chiong R, Bao Y, Hayat Malik B. Acceptance and use predictors of fitness wearable technology and intention to recommend. Industr Mngmnt Data Syst 2019 Feb 04;119(1):170-188. [doi: 10.1108/imds-01-2018-0009]
7. Gao Y, Li H, Luo Y. An empirical study of wearable technology acceptance in healthcare. Industr Mngmnt Data Syst 2015 Oct 19;115(9):1704-1723. [doi: 10.1108/IMDS-03-2015-0087]
8. Li J, Ma Q, Chan AH, Man SS. Health monitoring through wearable technologies for older adults: smart wearables acceptance model. Appl Ergon 2019 Feb;75:162-169. [doi: 10.1016/j.apergo.2018.10.006] [Medline: 30509522]

9. Zhang M, Luo M, Nie R, Zhang Y. Technical attributes, health attribute, consumer attributes and their roles in adoption intention of healthcare wearable technology. Int J Med Inform 2017 Dec;108:97-109. [doi: 10.1016/j.ijmedinf.2017.09.016] [Medline: 29132639]
10. Gao S, Zhang X, Peng S. Understanding the adoption of smart wearable devices to assist healthcare in China. In: Dwivedi Y, Mantymaki M, Ravishankar M, Janssen M, Clement M, Slade E, editors. Social Media: The Good, the Bad, and the Ugly. London: Springer; 2016:280-291.
11. Meng S, Hui L, peng W, Yong-ho S. Empirical research on the determinants of continuous use intention of users of smart health wearable device. J Math Pract Theory 2019;49(7):9-19.
12. Wiegard R, Guhr N, Krylow S, Breitner MH. Analysis of wearable technologies' usage for pay-as-you-live tariffs: recommendations for insurance companies. ZVersWiss 2019 Feb 18;108(1):63-88. [doi: 10.1007/s12297-019-00431-2]
13. Park E, Kim KJ, Kwon SJ. Understanding the emergence of wearable devices as next-generation tools for health communication. Inform Technol People 2016 Nov 07;29(4):717-732. [doi: 10.1108/itp-04-2015-0096]
14. Zhao Y, Ni Q, Zhou R. What factors influence the mobile health service adoption? A meta-analysis and the moderating role of age. Int J Inform Mngmnt 2018 Dec;43:342-350. [doi: 10.1016/j.ijinfomgt.2017.08.006]
15. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]
16. King WR, He J. A meta-analysis of the technology acceptance model. Inform Mngmnt 2006 Sep;43(6):740-755. [doi: 10.1016/j.im.2006.05.003]
17. Marangunić N, Granić A. Technology acceptance model: a literature review from 1986 to 2013. Univ Access Inf Soc 2014 Feb 16;14(1):81-95. [doi: 10.1007/s10209-014-0348-1]
18. Ooi K, Tan GW. Mobile technology acceptance model: an investigation using mobile users to explore smartphone credit card. Exp Syst Applic 2016 Oct;59:33-46. [doi: 10.1016/j.eswa.2016.04.015]
19. Wu J, Wang S. What drives mobile commerce? An empirical evaluation of the revised technology acceptance model. Inform Mngmnt 2005 Jul;42(5):719-729. [doi: 10.1016/j.im.2004.07.001]
20. Hsiao S, Tseng H. The impact of the moderating effect of psychological health status on nurse healthcare management information system usage intention. Healthcare (Basel) 2020 Feb 02;8(1):1 [FREE Full text] [doi: 10.3390/healthcare8010028] [Medline: 32024298]
21. Asadi S, Abdullah R, Safaei M, Nazir S. An integrated SEM-neural network approach for predicting determinants of adoption of wearable healthcare devices. Mobile Inform Syst 2019 Feb 14;2019:1-9. [doi: 10.1155/2019/8026042]
22. Taib S, De Coster R, Nyamu J. Innovation diffusion of wearable mobile computing: pervasive computing perspective. 2016 Presented at: International Conference on Information Society (i-Society); 2016; Dublin. [doi: 10.1109/i-society.2016.7854185]
23. Rajanen D, Weng M. Digitization for fun or reward? A study of acceptance of wearable devices for personal healthcare. Proc 21st Int Acad Mindtrek Conf 2017:1. [doi: 10.1145/3131085.3131118]
24. Baek M, Choi H, Lee H. Age-specific acceptance intention over wearable smart healthcare device. Korean J Bus Admin 2015 Dec 31;28(12):3171-3189. [doi: 10.18032/kaaba.2015.28.12.3171]
25. Choi B, Hwang S, Lee S. What drives construction workers' acceptance of wearable technologies in the workplace? Indoor localization and wearable health devices for occupational safety and health. Automation Construction 2017 Dec;84:31-41. [doi: 10.1016/j.autcon.2017.08.005]
26. Talukder MS, Sorwar G, Bao Y, Ahmed JU, Palash MAS. Predicting antecedents of wearable healthcare technology acceptance by elderly: a combined SEM-Neural Network approach. Technol Forecast Soc Change 2020 Jan;150:119793 [FREE Full text] [doi: 10.1016/j.techfore.2019.119793]
27. Yang Meier D, Barthelmess P, Sun W, Liberatore F. Wearable technology acceptance in health care based on national culture differences: cross-country analysis between Chinese and Swiss consumers. J Med Internet Res 2020 Oct 22;22(10):e18801 [FREE Full text] [doi: 10.2196/18801] [Medline: 33090108]
28. Rokeach M. Understanding Human Values. New York: Free Press; 1979.
29. Hanson GH. Market potential, increasing returns and geographic concentration. J Int Econ 2005 Sep;67(1):1-24. [doi: 10.1016/j.jinteco.2004.09.008]
30. Hofstede G. Cultures and Organizations: Software of the Mind. New York: Business Expert Press; 2010.
31. Hung C, Chou JC. Examining the cultural moderation on the acceptance of mobile commerce. Int J Innov Technol Mngmnt 2014 Mar 25;11(02):1450010. [doi: 10.1142/s0219877014500102]
32. Yuen Y, Yeow P, Lim N. Internet banking acceptance in the United States and Malaysia: a cross-cultural examination. Mark Intell Planning 2015;33(3):292-308. [doi: 10.1108/mip-08-2013-0126]
33. Lin H. An investigation of the effects of cultural differences on physicians' perceptions of information technology acceptance as they relate to knowledge management systems. Comput Human Behav 2014 Sep;38:368-380. [doi: 10.1016/j.chb.2014.05.001]
34. Baptista G, Oliveira T. Understanding mobile banking: the unified theory of acceptance and use of technology combined with cultural moderators. Comput Human Behav 2015 Sep;50:418-430. [doi: 10.1016/j.chb.2015.04.024]

35. Søndergaard M. Research note: Hofstede's consequences: a study of reviews, citations and replications. Org Stud 2016 Jun 30;15(3):447-456. [doi: 10.1177/017084069401500307]
36. Mooij M. Global Marketing and Advertising: Understanding Cultural Paradoxes. London: Sage Open; 2009.
37. Hofstede G. Dimensionalizing cultures: the Hofstede model in context. Online Readings Psychol Cult 2011 Dec 01;2(1):1-25. [doi: 10.9707/2307-0919.1014]
38. Aw EC, Chuah SH, Sabri MF, Kamal Basha N. Go loud or go home? How power distance belief influences the effect of brand prominence on luxury goods purchase intention. J Retailing Consumer Serv 2021 Jan;58:102288. [doi: 10.1016/j.jretconser.2020.102288]
39. Eastman JK, Iyer R, Shepherd CD, Heugel A, Faulk D. Do they shop to stand out or fit in? The luxury fashion purchase intentions of young adults. Psychol Mark 2018 Feb 09;35(3):220-236. [doi: 10.1002/mar.21082]
40. Veiga JF, Floyd S, Dechant K. Towards modelling the effects of national culture on it implementation and acceptance. J Inform Technol 2001 Sep;16(3):145-158. [doi: 10.1080/02683960110063654]
41. Gc V, Iglesias C, Erdem S, Hassan L, Manca A. PUK21 patients' preferences for a wearable digital health technology to support self-management of chronic kidney disease. Value Health 2021 Jun;24:S237. [doi: 10.1016/j.jval.2021.04.1189]
42. Dieffenderfer J, Goodell H, Mills S, McKnight M, Yao S, Lin F, et al. Low-power wearable systems for continuous monitoring of environment and health for chronic respiratory disease. IEEE J Biomed Health Inform 2016 Dec;20(5):1251-1264 [FREE Full text] [doi: 10.1109/JBHI.2016.2573286] [Medline: 27249840]
43. Carchiolo V, Longheu A, Tinella S, Ferrara S, Savalli N. Data collection via wearable medical devices for mobile health. 2020 Presented at: 13th International Joint Conference on Biomedical Engineering Systems and Technologies; 2020; Valletta. [doi: 10.5220/0009100705860592]
44. Cheung ML, Chau KY, Lam MHS, Tse G, Ho KY, Flint SW, et al. Examining consumers' adoption of wearable healthcare technology: the role of health attributes. Int J Environ Res Public Health 2019 Jun 26;16(13):2257 [FREE Full text] [doi: 10.3390/ijerph16132257] [Medline: 31247962]
45. Min Z, Meifen L, Rui N. Analysis on factors influencing consumer usage intention of healthcare wearable technology: based on the moderating effects of usage experience and health knowledge. J Inform Res Mngmnt (2) 2017:14-21. [doi: 10.13365/j.jirm.2017.02.014]
46. Venkatesh V, Morris M, Davis G, Davis F. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]
47. Hung Y, Yang H, Hsiao C, Yang Y. A study of behavioral intention for mobile commerce using technology acceptance model. 2004 Presented at: 4th International Conference on Electronic Business (ICEB 2004); 2004; Beijing.
48. Wu Z. Chinese customer's attitude and adopt intention on mobile commerce. 2009 Presented at: 6th International Conference on Service Systems and Service Management; 2009; Xiamen. [doi: 10.1109/icsssm.2009.5174977]
49. Hofstede G. Culture's Consequences:Comparing Values, Behaviors, Institutions and Organizations Across Nations. Thousand Oaks: Sage Publications, Inc; 2001.
50. Minkov M, Hofstede G. Hofstede's fifth dimension: new evidence from the world values survey. J Cross Cult Psychol 2010 Dec 15;43(1):3-14. [doi: 10.1177/0022022110388567]
51. Ge W, Bai H, Wu H, He J. Teacher supported behaviors affecting MOOC learners' intention based on TAM and SOR Model. 2020 Presented at: 9th Annual International Conference on Educational Innovation through Technology (EITT); 2020; Porto. [doi: 10.1109/eitt50754.2020.00013]
52. Hendrickson AR, Collins MR. An assessment of structure and causation of IS usage. The Data Base for Advances in Information Systems 1996 Apr;27(2):61-67. [doi: 10.1145/243350.243361]
53. Tarhini A, Hone K, Liu X, Tarhini T. Examining the moderating effect of individual-level cultural values on users' acceptance of E-learning in developing countries: a structural equation modeling of an extended technology acceptance model. Interact Learn Envir 2016 Jan 08;25(3):306-328. [doi: 10.1080/10494820.2015.1122635]
54. Zhu Y, Dailey SL, Kreitzberg D, Bernhardt J. "Social networkout": connecting social features of wearable fitness trackers with physical exercise. J Health Commun 2017 Dec;22(12):974-980. [doi: 10.1080/10810730.2017.1382617] [Medline: 29173072]
55. Shen J, Chen Y, Shen Z, Liu S. A two-stage incremental update method for fall detection with wearable device. 2018 Presented at: IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI); 2018; Guangzhou. [doi: 10.1109/smartworld.2018.00093]
56. Vu CC, Kim SJ, Kim J. Flexible wearable sensors: an update in view of touch-sensing. Sci Technol Adv Mater 2021 Mar 31;22(1):26-36 [FREE Full text] [doi: 10.1080/14686996.2020.1862629] [Medline: 33854405]
57. Azevedo D, Esteves A, Ribeiro F, Farinha L, Metrolho J. A wearable device for monitoring health risks when children play outdoors. 2020 Presented at: 15th Iberian Conference on Information Systems and Technologies (CISTI); 2020; Seville. [doi: 10.23919/cisti49556.2020.9140946]
58. Giansanti D, Morelli S, Maccioni G, Costantini G. Toward the design of a wearable system for fall-risk detection in telerehabilitation. Telemed J E Health 2009 Apr;15(3):296-299. [doi: 10.1089/tmj.2008.0106] [Medline: 19382869]

59. Minkov M. What Makes Us Different and Similar: A New Interpretation of the World Values Survey and Other Cross-cultural Data. Sofia: Klasika y Stil; 2007.
60. Cho I, Kaplanidou K, Sato S. Gamified wearable fitness tracker for physical activity: a comprehensive literature review. Sustainability 2021 Jun 22;13(13):7017. [doi: 10.3390/su13137017]
61. Jeong SC, Kim S, Park JY, Choi B. Domain-specific innovativeness and new product adoption: a case of wearable devices. Telematics Informatics 2017 Aug;34(5):399-412. [doi: 10.1016/j.tele.2016.09.001]
62. Cooper H. Research Synthesis and Meta-Analysis : A Step-By-Step Approach. Thousand Oaks: SAGE Publications; 2016.
63. Lin W, Ke H, Chou W, Chang P, Tsai T, Lee M. Realization and technology acceptance test of a wearable cardiac health monitoring and early warning system with multi-channel MCGs and ECG. Sensors (Basel) 2018 Oct 19;18(10):e18801 [FREE Full text] [doi: 10.3390/s18103538] [Medline: 30347695]
64. Chen X, Shen W. Empirical methods in organization and management research, the third edition. Peking University Press: Beijing; 2018.
65. Vallo Hult H, Hansson A, Gellerstedt M. Digitalization and physician learning: individual practice, organizational context, and social norm. J Contin Educ Health Prof 2020;40(4):220-227 [FREE Full text] [doi: 10.1097/CEH.0000000000000303] [Medline: 33284172]
66. Woo K, Dowding DW. Decision-making factors associated with telehealth adoption by patients with heart failure at home. Comput Informatics Nurs 2020;38(4):204-214. [doi: 10.1097/cin.0000000000000589]
67. Hoogenbosch B, Postma J, de Man-van Ginkel JM, Tiemessen NA, van Delden JJ, van Os-Medendorp H. Use and the users of a patient portal: cross-sectional study. J Med Internet Res 2018 Sep 17;20(9):e262 [FREE Full text] [doi: 10.2196/jmir.9418] [Medline: 30224334]
68. Beza E, Reidsma P, Poortvliet PM, Belay MM, Bijen BS, Kooistra L. Exploring farmers' intentions to adopt mobile Short Message Service (SMS) for citizen science in agriculture. Comput Electronics Agricult 2018 Aug;151:295-310. [doi: 10.1016/j.compag.2018.06.015]
69. Jadil Y, Rana NP, Dwivedi YK. A meta-analysis of the UTAUT model in the mobile banking literature: the moderating role of sample size and culture. J Bus Res 2021 Aug;132:354-372. [doi: 10.1016/j.jbusres.2021.04.052]
70. Hofstede Insights. URL: https://www.hofstede-insights.com/product/compare-countries/ [accessed 2021-09-04]
71. Light R, Pillemer D. Summing-Up: The Science of Reviewing Research. Cambridge: Harvard University Press; 1984:0362-9791.
72. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634 [FREE Full text] [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]
73. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 6;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
74. Schmidt FL, Oh I, Hayes TL. Fixed- versus random-effects models in meta-analysis: model properties and an empirical comparison of differences in results. Br J Math Stat Psychol 2009 Feb;62(Pt 1):97-128. [doi: 10.1348/000711007X255327] [Medline: 18001516]
75. Chang C. Exploring the usage intentions of wearable medical devices: a demonstration study. Interact J Med Res 2020 Sep 18;9(3):e19776 [FREE Full text] [doi: 10.2196/19776] [Medline: 32945778]
76. Chau KY, Lam MHS, Cheung ML, Tso EKH, Flint SW, Broom DR, et al. Smart technology for healthcare: exploring the antecedents of adoption intention of healthcare wearable technology. Health Psychol Res 2019 Mar 11;7(1):8099 [FREE Full text] [doi: 10.4081/hpr.2019.8099] [Medline: 31583292]
77. Jin S. A study on the effect of individual characteristics on acceptance intention of wearable healthcare devices: focusing on the UTAUT2 and innovativeness. J Korea Soc Comput Inform 2020;25(3):129-143. [doi: 10.9708/jksci.2020.25.03.129]
78. Kim J, Ri K. An analysis of cognitive ability and technology acceptance behavior for the elderly: towards the use of wearable healthcare devices. J Inform Technol Applic Mngmnt 2019;26(1):2019-2038. [doi: 10.21219/jitam.2019.26.1.021]
79. Kim TB, Ho CB. Validating the moderating role of age in multi-perspective acceptance model of wearable healthcare technology. Telematics Informatics 2021 Aug;61:101603. [doi: 10.1016/j.tele.2021.101603]
80. Jung M, Ha Y, Yang S. Factors influencing on workers' intention to use healthcare for wearable device. Asia Pacific J Multimedia Serv Convergent Art Humanities Sociol 2017 Jun 30;7(6):337-349. [doi: 10.14257/ajmahs.2017.06.03]
81. Lee S, Yoo W, Park H, Kim S. An empirical study on acceptance intention towards healthcare wearable device. J Inform Syst 2016 Jun 30;25(2):27-50. [doi: 10.5859/kais.2016.25.2.27]
82. Wang H, Tao D, Yu N, Qu X. Understanding consumer acceptance of healthcare wearable devices: an integrated model of UTAUT and TTF. Int J Med Inform 2020 Jul;139:104156. [doi: 10.1016/j.ijmedinf.2020.104156] [Medline: 32387819]
83. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale: L Erlbaum Associates; 1988.
84. Liu Z, Ben S, Zhang R. Factors affecting consumers' mobile payment behavior: a meta-analysis. Electron Commer Res 2019 Apr 24;19(3):575-601. [doi: 10.1007/s10660-019-09349-4]
85. Venkatesh V, Bala H. Technology acceptance model 3 and a research agenda on interventions. Decis Sci 2008 May;39(2):273-315. [doi: 10.1111/j.1540-5915.2008.00192.x]
86. Zhang Y, Weng Q, Zhu N. The relationships between electronic banking adoption and its antecedents: a meta-analytic study of the role of national culture. Int J Inform Mngmnt 2018 Jun;40:76-87. [doi: 10.1016/j.ijinfomgt.2018.01.015]

87. Humana C. World Human Rights Guide. London: Oxford University Press; 1992.
88. Yoon C. The effects of national culture values on consumer acceptance of e-commerce: online shoppers in China. Inform Mngmnt 2009 Jun;46(5):294-301. [doi: 10.1016/j.im.2009.06.001]

Abbreviations

TAM: technology acceptance model
UTAUT: unified theory of acceptance and use of technology

Edited by L Buis; submitted 04.06.21; peer-reviewed by D Yang Meier, S Rostam Niakan Kalhori, D Frost; comments to author 11.11.21; revised version received 05.01.22; accepted 09.02.22; published 03.06.22.

Please cite as:
Zhang Z, Xia E, Huang J
Impact of the Moderating Effect of National Culture on Adoption Intention in Wearable Health Care Devices: Meta-analysis
JMIR Mhealth Uhealth 2022;10(6):e30960
URL: https://mhealth.jmir.org/2022/6/e30960
doi:10.2196/30960
PMID:35657654

Review

Emerging Artificial Intelligence–Empowered mHealth: Scoping Review

Paras Bhatt[1], MSc; Jia Liu[2], PhD; Yanmin Gong[1], PhD; Jing Wang[3], PhD; Yuanxiong Guo[1], PhD

[1]Department of Electrical & Computer Engineering, The University of Texas at San Antonio, San Antonio, TX, United States

[2]The University of Texas Health Science Center at San Antonio, San Antonio, TX, United States

[3]Florida State University, Tallahassee, FL, United States

Corresponding Author:
Yuanxiong Guo, PhD
Department of Electrical & Computer Engineering
The University of Texas at San Antonio
1 UTSA Circle
San Antonio, TX, 78249
United States
Phone: 1 210 458 8028
Email: yuanxiong.guo@utsa.edu

Abstract

Background: Artificial intelligence (AI) has revolutionized health care delivery in recent years. There is an increase in research for advanced AI techniques, such as deep learning, to build predictive models for the early detection of diseases. Such predictive models leverage mobile health (mHealth) data from wearable sensors and smartphones to discover novel ways for detecting and managing chronic diseases and mental health conditions.

Objective: Currently, little is known about the use of AI-powered mHealth (AIM) settings. Therefore, this scoping review aims to map current research on the emerging use of AIM for managing diseases and promoting health. Our objective is to synthesize research in AIM models that have increasingly been used for health care delivery in the last 2 years.

Methods: Using Arksey and O'Malley's 5-point framework for conducting scoping reviews, we reviewed AIM literature from the past 2 years in the fields of biomedical technology, AI, and information systems. We searched 3 databases, PubsOnline at *INFORMS*, e-journal archive at *MIS Quarterly*, and Association for Computing Machinery (ACM) Digital Library using keywords such as "mobile healthcare," "wearable medical sensors," "smartphones", and "AI." We included AIM articles and excluded technical articles focused only on AI models. We also used the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) technique for identifying articles that represent a comprehensive view of current research in the AIM domain.

Results: We screened 108 articles focusing on developing AIM models for ensuring better health care delivery, detecting diseases early, and diagnosing chronic health conditions, and 37 articles were eligible for inclusion, with 31 of the 37 articles being published last year (76%). Of the included articles, 9 studied AI models to detect serious mental health issues, such as depression and suicidal tendencies, and chronic health conditions, such as sleep apnea and diabetes. Several articles discussed the application of AIM models for remote patient monitoring and disease management. The considered primary health concerns belonged to 3 categories: mental health, physical health, and health promotion and wellness. Moreover, 14 of the 37 articles used AIM applications to research physical health, representing 38% of the total studies. Finally, 28 out of the 37 (76%) studies used proprietary data sets rather than public data sets. We found a lack of research in addressing chronic mental health issues and a lack of publicly available data sets for AIM research.

Conclusions: The application of AIM models for disease detection and management is a growing research domain. These models provide accurate predictions for enabling preventive care on a broader scale in the health care domain. Given the ever-increasing need for remote disease management during the pandemic, recent AI techniques, such as federated learning and explainable AI, can act as a catalyst for increasing the adoption of AIM and enabling secure data sharing across the health care industry.

(JMIR Mhealth Uhealth 2022;10(6):e35053) doi:10.2196/35053

KEYWORDS

mobile health units; telemedicine; machine learning; artificial intelligence; review literature as topic

Introduction

Initially, information technology systems were mainly used to record patient data [1], but the rapid development in technology over the years has paved the way for data analytics and machine learning (ML) to be applied in the health care domain [2]. Advanced artificial intelligence (AI) techniques combined with the rapid integration of medical internet of things (IoT) devices [3] has led to an increase in research on digital health care and preventive medicine [4]. Such research focuses on mobile health (mHealth) technologies that are used to monitor serious ailments, like asthma, diabetes, and sleep apnea, and to ensure patient well-being and safety [5]. mHealth is a critical sector of the health care information technology industry that has grown rapidly in recent years [6]. This growth has been fueled by the rise in wearable technologies [7], mobile sensors [8], and the exponential increase in the number of IoT devices in general [9]. Such devices are increasingly used in hospitals and medical institutions [10] for constant patient monitoring [11] and intensive care unit capacity monitoring. Coinciding with the increase in the use of IoT devices, the wearing of health devices outside hospital premises for remote in-home care has also increased [12]. This has led to both a greater level of research [13] and higher investment in mHealth [14]. Researchers have stressed mHealth's importance in challenging times such as the current pandemic to enable the provision of remote health care facilities [15]. Recent research indicates that there has been a significant increase in mHealth usage since COVID-19 [16]. AI has helped scholars to research new avenues of clinical care that are focused on ensuring the maintenance of social distancing and better hygiene and have developed remote mHealth capabilities that can enable patient care during and after COVID-19 [17,18].

With the increase in mHealth research, there have been significant improvements in the level of AI available to researchers as well. These improvements offer more accurate insights and results than does traditional ML while simultaneously preserving patient privacy and ensuring a high data security standard. Deep learning (DL) and federated learning (FL) [19,20] are some examples of these newer techniques that ensure data security and privacy. Consequently, researchers have used AI techniques to study novel scenarios and tasks within the health care IT domain, from using it to classify and predict disease occurrence [21], to detecting the presence of chronic illnesses [22], and even assisting doctors in making decisions about preventive health care programs [23]. AI has been successfully integrated with the health care sector, and many systematic literature surveys outline its importance to this domain [24-26].

Recently, research related to AI techniques in the mHealth sector has increased considerably [27]. This can be attributed to the rapid evolution and acceptance of telehealth during the COVID-19 pandemic [28]. As a result of several changes in telehealth policies (telehealth integration into hospital portals, expanding insurance coverage for telehealth services, and increasing patient choices for telehealth services) [29,30], telehealth has emerged as a viable alternative to providing care to noncritical patients [31], thus enabling hospitals and medical institutions to direct their resources to serving critical patients. The adoption of mHealth devices has also increased during this period [16], providing both localized and personalized patient information [32] and resulting in the generation of a large amount of data which is particularly well-suited to train AI models. ML algorithms running locally on smart and wearable devices have led to novel insights. For instance, researchers use AI to study neurogenerative disorders like Parkinson [33] and Alzheimer [34] disease, which exhibit latent temporal symptoms that are difficult to characterize without mHealth sensors. This symbiosis of AI and mHealth technologies is crucial for the development of remote health care infrastructure that can better inform physicians and benefit millions of patients.

Using mHealth sensors, researchers have documented disease progression [35], depicting how an illness spreads or manifests over time in a patient. These insights can be significant in the early diagnosis and treatment of chronic diseases and management of symptoms hitherto undetectable by traditional patient monitoring within hospitals and assisted living facilities. This confluence of AI and mHealth has given rise to a new domain of research that studies the combination of these 2 research streams. It is known as AI-powered mHealth (AIM) [36]. Using AI techniques in the application of mHealth scenarios can have numerous benefits, such as the automatic detection of chronic disease occurrence [21], real-time prediction and intervention for suicide prevention [37], facilitating emergency response [38], enabling patient rehabilitation [39], providing noninvasive care [40,41], and preventing medical errors. Preventable medical error is a significant cause of death in the USA. Clinical decision-making technology can significantly reduce it by using real-time data from wearable health sensors [42,43].

AIM devices can power ubiquitous health care solutions through remote patient monitoring [44], which is essential for providing health services in remote and medically underserved areas, where patients are not connected with modern health care systems. AIM can also enable at-risk minority populations who do not have access to health care facilities receive quality health care with ease [45,46]. With the development of newer AI techniques, such as DL, reinforcement learning, and few-shot learning, the domain of AIM will only grow in the future [47,48]. Furthermore, mHealth has implications for remote patient-monitoring and telehealth research and practice, which is becoming a reality much faster than the medical industry expected because of the COVID-19 pandemic [49].

Prior research focusing on the application of AI in the health care domain noted that certain implementation factors exist that prevent large scale automation of the health care sector [50]. However, with the advancement in AI techniques and the advent of DL, there has been a significant rise in both AIM research and practice. Previous surveys of mHealth have focused on only

niche conditions, such as musculoskeletal medicine [51], or have attempted to study perceptions of AI in the mHealth domain and health care settings in general [52,53].

Over the last couple of years, there have been significant advances in both the usage of AI and mHealth. In this regard, several recent studies share an overlapping context (AI + mHealth) [54-57] in seeking to explain and implement the clinical use of AI in mHealth settings. A current review of such research is lacking, which presents a gap in the AI mHealth literature. Therefore, it is necessary to survey the current state of the art in AIM research (eg, current work, current solutions, and future opportunities) both in the mHealth industry and the field of AI. A scoping review of this research is much needed, as it addresses the gap in literature related to an in-depth analysis of AI capabilities currently being used in the mHealth settings. Our aim in this paper is to further expand the research scope of this critical health care domain and explore the opportunities for future development. To the best of our knowledge, this is the first attempt to survey research on AIM analytics. Our objective is also to map current research on the growing use of AIM for remote patient-monitoring and examine how researchers use patient data for building AI models for disease management.

Methods

Scoping reviews are used to examine the extent, range, and nature of research activity in a particular domain. In this context, we used Arksey and O'Malley's [58] 5-step scoping review framework to guide our search strategy for reviewing current peer-reviewed AIM research.

Step 1: Identifying Research Questions

We started by identifying our research questions (RQs) and aimed to survey the literature on the current use of AIM to identify and manage different health conditions. We also investigated the use of data collected from wearable sensors and mobile devices for building AIM models.

Step 2: Identifying Relevant Studies

After specifying our RQs, we identified relevant studies to be screened in this review. This involved searching electronic databases including PubsOnline at *INFORMS* and the e-journal archive at *MIS Quarterly* for information systems (IS) articles. We used the Association for Computing Machinery (ACM) Digital Library, which catalogs research from top conferences and journals in the AI domain, for AI articles. We also used a search query in Google Scholar with a 2-year filter (since 2019) for including recent articles on specific advancements in the field of AIM related to the use of FL and explainable AI (XAI). The state of the art in AI until 2019 had been covered by previous researchers in surveys and reviews on AI in the health care sector [50,51,53]; therefore, we decided to focus on articles from 2019 and beyond. Moreover, since DL has only been growing in the health care domain during the last couple of years and newer AIM techniques such as FL have recently emerged as privacy-preserving mechanisms, we decided to limit the search to articles published from 2019 to the present.

The articles screened for this review were published in the 3 major domains of biomedical technology, AI, and IS. In this regard, the journals and articles searched were from top venues in these domains. We searched the *Journal of Biomedical Informatics*, *Journal of Medical Internet Research*, and *Nature Medicine* for biomedical technology articles. For AI, we focused our efforts on recent top conferences including the Conference on Neural Information Processing Systems (NeurIPS) and the Association for the Advancement of Artificial Intelligence (AAAI). Within these conferences, we looked at ML for health, ML for mobile health, ML for public health, and web search and data mining. As for IS, we searched articles in the top journals of *Management Information Systems Quarterly* (*MISQ*), *Information Systems Research* (*ISR*), and ACM's Transactions journals. These studies are related to the use of health care technology combined with a behavioral component that seeks to explain how AI can define patient well-being. We used the keywords "mobile health," "mHealth," "mobile healthcare," "mobile sensors," "wearable sensors," "medical sensors," "smartphone data," "ML," and "AI."

Step 3: Study Selection

After selecting relevant articles, we defined our study selection metrics based on the inclusion and exclusion criteria as specified in Table S1 in Multimedia Appendix 1. We eliminated several articles identified through our keyword searches that did not meet the criteria. During the last 2 years, there has been a significant increase in research in AIM [59]. In the same period, researchers have developed and used newer AI techniques, such as FL and XAI, to build predictive privacy-preserving models [20] for disease management. Therefore, we decided to limit the search for AIM articles to the last 2 years. Additionally, we considered articles where both AI and mHealth concepts were specifically used in the study design or the primary research motivation for the paper. Finally, each author independently read article summaries and abstracts to determine their eligibility for this scoping review based on the inclusion and exclusion criteria as specified in Table S1 in Multimedia Appendix 1.

Step 4: Charting the Data

After selection of the studies for this scoping review, we segregated articles according to research streams (biomedical technology, AI, IS), type of data used (public vs proprietary), and health conditions (physical health, mental health, and general health promotion and wellness). Articles related to physical health examined the use of AIM for disease management of chronic health issues, such as asthma and diabetes, and neurological illnesses, such as Alzheimer and Parkinson disease. These severe health conditions are difficult to manage without active support from physicians, and thus, the application of mHealth sensors can be used to track patients with these conditions. Studies focused on general health promotion and wellness were related to nonchronic conditions that do not require constant medical supervision, such as leading an active lifestyle and engaging in regular exercise. mHealth sensors can notify and remind people to engage in physical activity to have an overall better level of physical health. Articles related to mental health focused on using AIM to facilitate the

detection of mental health issues among the population by collecting data from personal devices.

Since public data has massive potential for enabling broader collaboration in health care technology usage and AIM adoption across organizational, national, and international boundaries, we also divided articles based on the data set they used. Studies using publicly available data sets are more effective in bringing out the potential impact of AIM and inspiring confidence among the public and medical institutions in the efficacy of AIM models.

Step 5: Collating, Summarizing, and Reporting the Results

Finally, in the results section, we collate, summarize, and report the findings from this review. We discuss their implications for future AIM research in the discussion section and present results of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) technique we used to identify and select articles for this review. We also discuss the selected 37 articles using AIM capabilities for disease management and monitoring physical and mental health conditions. Furthermore, some of the articles focused on using AIM models for enhancing general health promotion and wellness of people.

Results

Step 1: Identifying RQs

After careful consideration and discussions, we decided to define the scope of our review paper based on the shared capabilities of AI and mHealth. Through deliberations, we decided to focus on the emerging uses of AI in the current state-of-the-art mobile health care domain. In this regard, we identified the following 3 RQs of value for both researchers and practitioners in this scoping review: (1) What are the major health conditions being researched in the AI-powered mHealth (AIM) domain? (2) How do AIM techniques use the data collected from wearable sensors and mHealth devices? (3) What are the requirements for facilitating the rapid adoption of AIM models in hospitals and medical institutions in the health care sector?

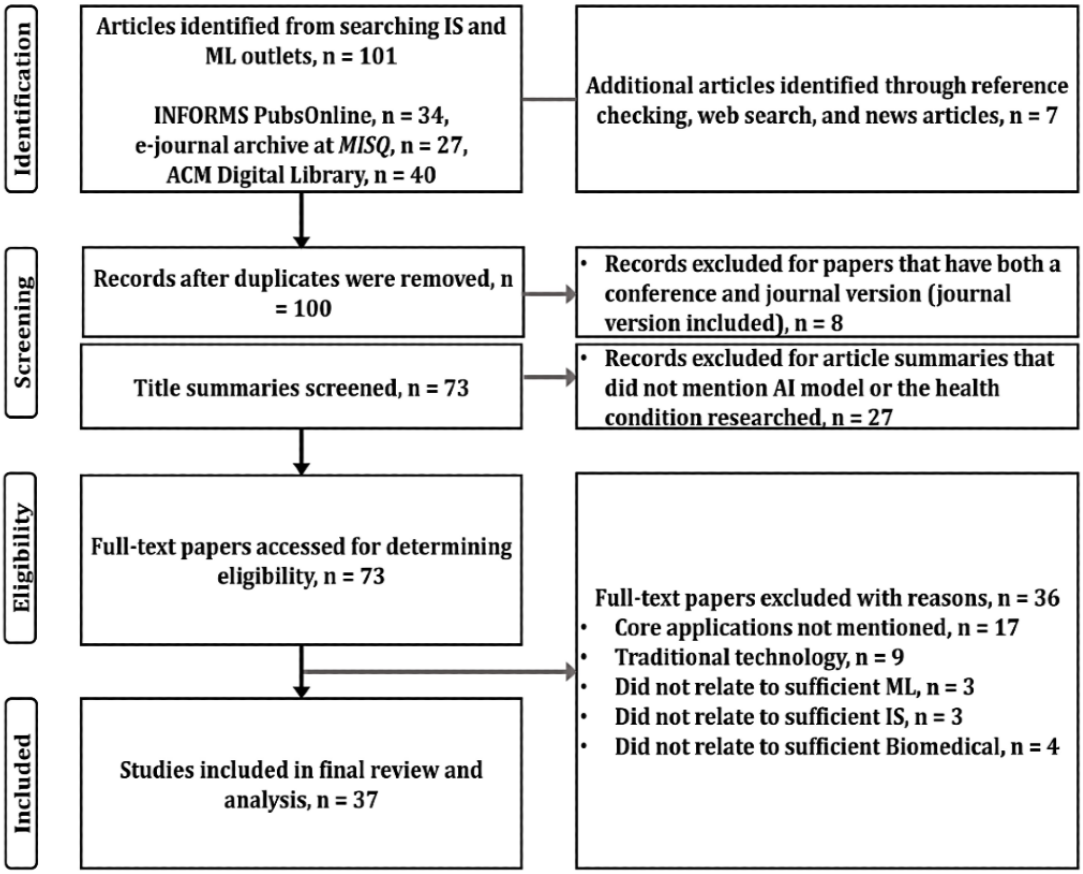
Step 2: Identifying Relevant Studies

We initially started with 108 articles related to each of the 3 domains in this study: biomedical technology, AI, and IS. We identified 108 relevant studies in total: 101 from our selected databases (PubsOnline, n=34; e-journal archive at *MISQ*, n=27; and ACM Digital Library, n=40) and 7 articles through reference checking in search engines.

Step 3: Study Selection

Using the PRISMA technique depicted in Figure 1, 37 articles matched the study selection criteria for this scoping review. When selecting the articles, we proceeded to remove duplicate articles (n=8) that had both a journal and conference version (journal version included in review) and screened the title and abstract of the selected articles (n=27) to ensure sufficient AI- and mHealth-based content was present in the work. Upon final selection, we independently screened the full text of the remaining articles (n=3).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram. ACM: Association for Computing Machinery; AI: artificial intelligence; IS: information systems; MISQ: Management Information Systems Quarterly; ML: machine learning.

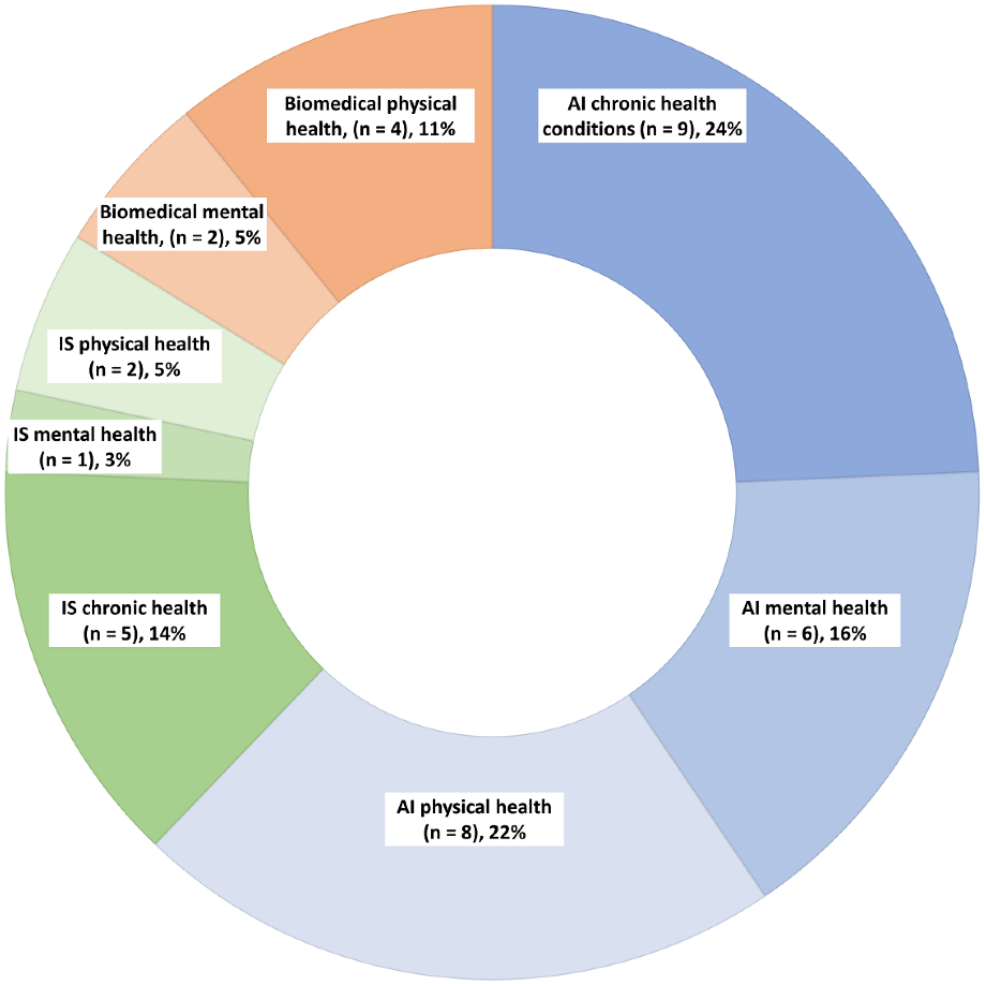


Step 4: Charting the Data

The majority of the articles identified used mixed method research with participant-based studies that focused on using mHealth devices to collect data from people experiencing a certain health condition (asthma, diabetes, suicidal tendencies, depression, etc) and then using the data collected to train AI models to automatically detect such conditions; otherwise, they were analytical studies that applied AI models to publicly available data sets. The final set of 37 articles included in this review are presented in Table S2 in Multimedia Appendix 1.

Of the 37 articles, 31 were published in 2020 (84%), and 23 out of the 37 (62%) articles were from AI databases which represented the largest domain included in our review. Of the 23 articles within the AI domain, 17 (74%) mainly focused on physical health and chronic health conditions. Both these conditions were researched in most of the articles included in the review. Physical health articles primarily focused on using AIM models for human activity recognition and analyzing people's activities of daily living. The data used in these studies were collected using multiple mHealth devices, such as object and motion detection sensors. However, accessing large repositories of such data is difficult because data sharing among medical institutions, hospitals, and clinical studies is often restricted [20]. This gives rise to a lack of availability of quality data sets for building AIM models, which was also observed in our review, as 28 out of the total 37 (76%) articles used proprietary data sets rather than public ones. Finally, mental health studies used a combination of qualitative techniques, such as surveys and smartphone sensors, to augment their data collection. These data were used for building predictive models that detect depression and suicidal tendencies in people. Figure 2 below presents the different metrics of the selected articles in our scoping review.

Figure 2. Statistics of different AI-powered mobile health domains (N=37). AI: artificial intelligence; IS: information systems.



Step 5: Collating, Summarizing, and Reporting the Results

The results from our scoping review help to answer our RQ1 about the major health conditions being researched in the AIM domain and our RQ2 about how AIM data collected from wearable sensors and mHealth devices are used by researchers. In terms of RQ1, 3 major categories are being researched in the AIM domain: mental health, physical health, and chronic health conditions. For RQ2, most of the studies use data collected from AIM devices to build and train advanced AI models that seek to detect, predict, and manage health conditions in general.

As we discussed the results of our scoping review of articles on AIM, we observed the research in this domain is concentrated in 3 distinct categories of physical health, mental health, and chronic health conditions, as presented in Figure 3.

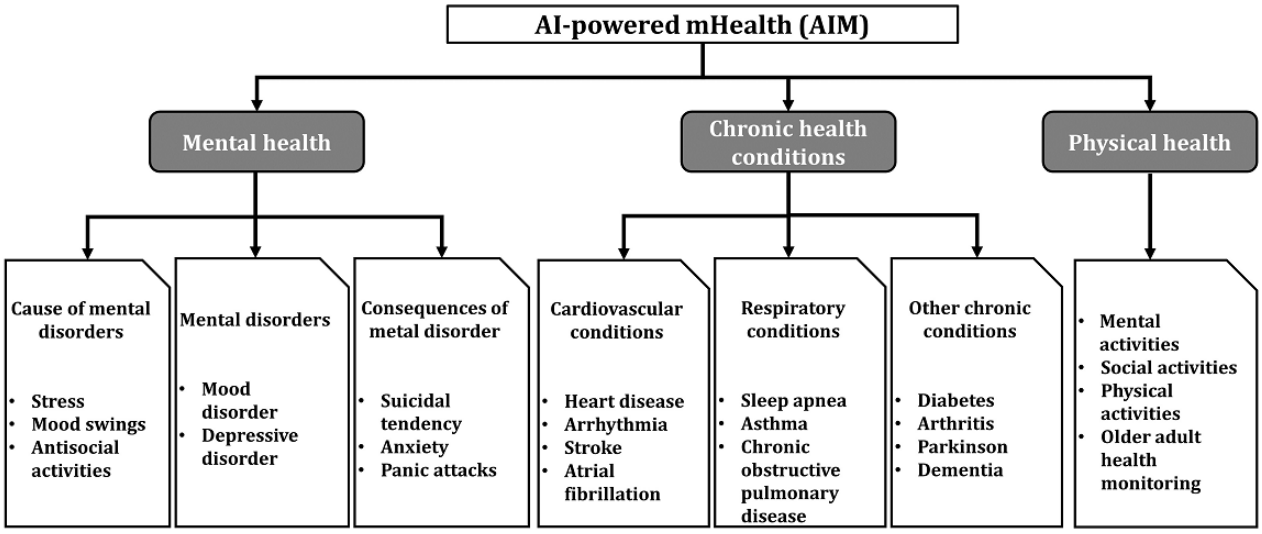
Most of the studies in the scoping review focused on chronic health conditions, such as cardiovascular conditions related to heart disease, stroke, arrhythmia, and atrial fibrillation; respiratory conditions, such as sleep apnea, asthma, and COVID-19 monitoring; and other conditions chronic conditions, such as diabetes, arthritis, and Parkinson disease. These studies explain how AIM models are used to enable greater self-management of chronic diseases by providing real-time health insights to patients and doctors [56,60]. AIM models focused on chronic health conditions are developed using heterogeneous data, including text, audio, and rhythmic body movements, collected from wearable and mobile sensors, [54,55,57,61]. Researchers note that the physiological features of people, such as their height, weight, and metabolism, can be used as data points to train personalized AIM models [62]. These models can then predict the types of chronic health conditions a person may be susceptible to (currently and in the future) [63]. For instance, researchers used AIM models to predict the likelihood of an imminent episode of Parkinson that may result in a patient falling [64]. Moreover, other researchers have demonstrated the effectiveness of AIM models in enhancing the development of preventive and precision medicine and detecting early signs of the onset of chronic conditions, such as in imminent asthma attacks [65,66].

After chronic conditions, the next major category of studies focused on mental health conditions for which research has recently increased [67]. Included were articles that sought to understand the nature, causes, and consequences of mental disorders. Studies in this category focused on using mobile sensors to predict people's moods and behaviors while also determining the causes responsible for such a shift in them [68,69]. Research aimed at understanding the causes and consequences of mental disorders is fundamental in clinical psychiatry [70] and can be used to provide interventions to people who exhibit antisocial behavior [71]. Studies relating to the consequences of mental health disorders play a crucial role

in identifying at-risk populations who might be suffering from suicidal tendencies. In some of our selected articles, researchers inferred suicidal tendencies from smartphone usage [72,73]. Smartphone usage can also help the understanding of the causes of mental health disorders. The mood of a person is indicative of the emotional state they are in. There is growing evidence to suggest people's moods (happy, sad, etc) and their inner emotional states (anxiety, depression, panic, etc) are interlinked [74,75]. Several studies have successfully pursued this link to identify people vulnerable to suffering from mental disorders [76,77]. For instance, some authors [76] used an AIM model to find emotionally distressed people in online social networks. The model analyzes the text of users' posts to detect the usage of negative words or phrases (eg, "I am all alone," "I don't want to live anymore") that indicate if a person is feeling suicidal. Similarly, other authors [77] also used data from wearable sensors to build AIM models that detect whether people are under emotional stress and determine the underlying causes for mental disorders.

The final category of studies focused on the use of AIM devices to monitor people's physical health. The articles studied different mental, social, and physical activities that people engage in and collected data using AIM devices. These data were used to build AIM models that detect when people were not engaging in their regular activities, such as exercising and walking. Once a lack of social or physical activity is detected, the AIM model sends out personalized suggestions encouraging people to lead an active lifestyle [78,79]. Such AIM models can also detect prolonged periods of human inactivity, which is of particular importance when monitoring the health of older adults. Studies show data related to heart rate and self-reported fatigue levels can be used to share automatic suggestions that remind people to engage in healthy exercise [80-83]. In addition, AIM devices can be used for monitoring the movements of older adults through the use of mobile sensors, such as wearable and object detection sensors [84-87]. Data collected through AIM devices can also identify human activity and encourage safe physical health practices [59,88]. For instance, during the current pandemic, researchers have used AIM devices to build models that identify and detect dangerous COVID-19 behavior, such as face touching [89,90]. Further, the use of AIM devices can help to ensure privacy and protect people's private health data [91] by using AI techniques, such as FL, that can prevent data from being transferred outside AIM devices. These kinds of varied applications showcase the versatility of the AIM domain for ensuring the physical health of people.

Figure 3. Studies in the AIM Domain. mHealth: mobile health.



Discussion

The findings from our scoping review showcase the benefits of using AIM in applications ranging from clinical care [92] to improvements in the overall adoption and access of telehealth services [29,30]. A key finding to discuss from our work is how the recent confluence of AI and mobile wearable technology has resulted in the increase in mHealth usage [16]. Another important insight to consider is how mHealth and telehealth have emerged as reliable avenues to provide noncritical care to patients [31], which is vital during the current pandemic. This is also evident in many of the papers included in this review [78,81,83].

This review surveys the recent developments in the AIM domain, and based on our findings, we present some practical recommendations for future research. First, by using the recent advances in AI techniques of FL and XAI, AIM could facilitate an even broader expansion and provision of mHealth services which is also evidenced by its significant use during the pandemic [16]. Second, the increased adoption of such services can be helpful in building a consensus about the rules governing the usage of wearable technologies in medical institutions. Currently, hospitals use proprietary mHealth devices that do not allow data sharing even if it is for critical research purposes [20]. With the increase in adoption, different health care institutions can come together and create a shared set of rules that can enable AI-based models to study the data from across all participating institutions, thus resulting in the generation of more robust and accurate medical insights. Third, given the importance of health care access for all sections of society, governments and private institutions should promote the use of

mHealth in their digital health efforts and public safety campaigns.

Given the growing importance of remote health and telehealth facilities during the COVID-19 pandemic, we conclude that it is important to facilitate greater adoption of remote health monitoring devices, protect patient privacy, and increase people's trust in smart wearable health devices. This is because our findings show that mHealth indeed plays a significant role in shaping the future of how citizens access health care facilities in testing times such as a pandemic. In this regard, some recent AI techniques that can accelerate its adoption and enable faster implementation across hospitals, medical institutions, and users at large are discussed. The discussion of these techniques also answers RQ3 about the requirements for facilitating the rapid adoption of AIM models in hospitals and medical institutions in the health care sector and how recent AI techniques can be implemented to strengthen research in the AIM domain.

To protect patients' privacy, ML techniques such as DL, FL, and transfer learning can effectively drive the smart health care revolution. These techniques use privacy-preserving feature engineering to translate vast amounts of biomedical data into actionable and potentially life-saving human health outcomes [93]. From the analysis of the papers included in this survey, we observe that a critical outcome of applying DL in the mHealth analytics domain is that it results in the development of powerful algorithms. These algorithms provide excellent capabilities to predict and detect diseases early, thus enabling efforts to provide preventive medicine and care to vulnerable people [94]. Since users' data exist in isolated silos or islands across different hospitals and medical institutions, it becomes increasingly difficult for researchers working in this domain to access these data. Moreover, generalizing the performance of an ML model for a large population becomes difficult in the absence of personalized data about individuals [95]. Recent advances in FL and transfer learning show that it is a promising solution in such scenarios. It ensures data privacy, as user data never leave an institution [96]. In addition, model insights learned from one set of data can be transferred to make predictions for another set of data. When FL models are used, data remain static and situated at the source, thereby protecting privacy. The only information exchange under such models involves purely numerical representations of stochastic gradient descent. These numerical data cannot be used to reverse engineer and determine the source of data. The use of such techniques can help dismantle privacy barriers that are associated with health care data access. It can enable greater collaboration between the medical, research, and practitioner communities while ensuring faster development and integration of AIM in the health care sector. To this effect, the 3 identified research streams can act as guiding principles for providing holistic health care services that cover the mind, body, and spirit of people. It can also ensure that people receive the best possible care in the shortest time and with maximum efficacy.

The DARPA (Defense Advanced Research Projects Agency) XAI program strives to support the development of AI systems whose models can be interpreted, understood, and trusted by end users [97]. XAI is necessary for the future of AIM integration, as it can help increase the acceptability and understanding of ML techniques and models in the health care domain. With increased knowledge of AI models, we can expect an increase in the adoption rate of AIM in the health care industry, as is evidenced by various studies relating to the technology acceptance model [98]. According to this model, as the ease of use of technology increases, so does the intention and behavior of people to engage in and use the said technology. In this case, as AI models become increasingly easy to use and deploy, their widespread adoption will significantly increase hospitals' efficacy. It will also result in better providing remote health services that depend on crucial data from patients' wearable sensor devices. Adadi et al [99] conducted in-depth survey on XAI and note its diverse implications for the medical field in the future. They emphasize how the lack of transparency in ML models is one of the primary reasons for the nonadoption of AI in the health care industry. Peeking inside the black-box nature of AI is thus an effective way to overcome the impediments that limited knowledge and understanding place on the use of AIM. Gordon et al [100] have shown how XAI techniques can be used in surgical and operative settings in hospitals and in processing medical data for real-time clinical decision support. These models can help surgical teams to analyze, anticipate, understand, and prevent adverse intraoperative events. In another study, Payrovnaziri et al [101] surveyed how XAI specifically can be used to model real-world electronic health record data. They identify several gaps in the literature and conclude XAI has not been adequately pursued and practiced in medicine. They acknowledge there are several opportunities available where the adoption and application of XAI can significantly enhance mHealth. These have important implications for both research and practice. The recency of these surveys underscores the importance of AIM in the health care sector and provides a guideline for future research into this critical domain.

As with most scoping reviews, there are some limitations in this work. First, we only considered research from the 3 domains of biomedical technology, IS, and AI. Second, we did not consider the social aspect of AIM technology in this paper, but it is an emerging aspect of health care research. We will work to address these limitations in our future work. Third, we considered only a limited number of databases for selecting the articles and had to restrict the search so that we could focus on articles that address the latest transdisciplinary research context of AI, biomedical technology, and IS. Such work included papers that were published in niche ML and AI conference proceedings and listed within a particular database, for instance the ACM Digital Library. However, the databases we selected are comprehensive avenues for state-of-the-art research in the AIM domain and include the latest peer-reviewed research literature in the 3 streams.

Our findings from this scoping review indicate that there has recently been considerable increase in the research, practice, and adoption of mHealth and AI capabilities in the health care sector, which has resulted in significant advances in both critical and noncritical clinical care. However, certain areas still exist where there is a lack of AI research, such as in addressing mental health issues. A particular reason for this lack of research can be attributed to the nonavailability of public data sets hindering

the widespread adoption of the AIM domain. A solution for this problem is to ensure collaboration and data sharing among different medical institutions. Such collaborative efforts will ensure the better utilization of AI tools by doctors, physicians, and hospitals alike. Furthermore, new and advanced AI techniques, such as FL and XAI, are rapidly being developed by researchers, and their subsequent adoption in real-world scenarios will likely have life-saving consequences in the future.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Selection criteria and study details of the artificial intelligence–powered mobile health (AIM) articles in the scoping review.
[DOCX File , 40 KB - mhealth_v10i6e35053_app1.docx]

References

1. Parente ST, McCullough JS. Health information technology and patient safety: evidence from panel data. Health Aff (Millwood) 2009;28(2):357-360. [doi: 10.1377/hlthaff.28.2.357] [Medline: 19275990]
2. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2014;2:3 [FREE Full text] [doi: 10.1186/2047-2501-2-3] [Medline: 25825667]
3. Valanarasu R. Smart and secure IoT and AI integration framework for hospital environment. JISMAC 2019 Nov;01(03):172-179. [doi: 10.36548/jismac.2019.3.004]
4. Barrett M, Boyne J, Brandts J, Brunner-La Rocca H, De Maesschalck L, De Wit K, et al. Artificial intelligence supported patient self-care in chronic heart failure: a paradigm shift from reactive to predictive, preventive and personalised care. EPMA J 2019 Dec;10(4):445-464 [FREE Full text] [doi: 10.1007/s13167-019-00188-9] [Medline: 31832118]
5. Guillodo E, Lemey C, Simonnet M, Walter M, Baca-García E, Masetti V, HUGOPSY Network, et al. Clinical applications of mobile health wearable-based sleep monitoring: systematic review. JMIR Mhealth Uhealth 2020 Apr 01;8(4):e10733 [FREE Full text] [doi: 10.2196/10733] [Medline: 32234707]
6. Chib A, Lin SH. Theoretical advancements in mHealth: a systematic review of mobile apps. J Health Commun 2018 Nov 19;23(10-11):909-955. [doi: 10.1080/10810730.2018.1544676] [Medline: 30449261]
7. Kalantari M. Consumers' adoption of wearable technologies: literature review, synthesis, and future research agenda. IJTMKT 2017;12(3):274-307. [doi: 10.1504/ijtmkt.2017.089665]
8. Tang T, Ho AT. A path-dependence perspective on the adoption of Internet of Things: Evidence from early adopters of smart and connected sensors in the United States. Government Information Quarterly 2019 Apr;36(2):321-332. [doi: 10.1016/j.giq.2018.09.010]
9. Singh D, Tripathi G, Jara A. A survey of Internet-of-Things: future vision, architecture, challenges and services. 2014 IEEE World Forum on Internet of Things (WF-IoT). Published online March 2014:287-292. [doi: 10.1109/wf-iot.2014.6803174]
10. Hsu C, Lin JC. An empirical examination of consumer adoption of Internet of Things services: network externalities and concern for information privacy perspectives. Computers in Human Behavior 2016 Sep;62:516-527. [doi: 10.1016/j.chb.2016.04.023]
11. Gómez J, Oviedo B, Zhuma E. Patient monitoring system based on internet of things. Procedia Computer Science 2016;83:90-97. [doi: 10.1016/j.procs.2016.04.103]
12. Uddin M, Alam J, Banu S. Real time patient monitoring system based on internet of things. 2017 Presented at: 4th International Conference on Advances in Electrical Engineering (ICAEE); Sep 28-30; Bangladesh. [doi: 10.1109/icaee.2017.8255410]
13. Olla P, Shimskey C. mHealth taxonomy: a literature survey of mobile health applications. Health Technol 2015 Jan 30;4(4):299-308. [doi: 10.1007/s12553-014-0093-8]
14. Malvey DM, Slovensky DJ. Global mHealth policy arena: status check and future directions. mHealth 2017 Sep 22;3:41-41. [doi: 10.21037/mhealth.2017.09.03]
15. 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]
16. 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]
17. Bhaskar S, Bradley S, Sakhamuri S, Moguilner S, Chattu VK, Pandya S, et al. Designing Futuristic Telemedicine Using Artificial Intelligence and Robotics in the COVID-19 Era. Front Public Health 2020;8:556789 [FREE Full text] [doi: 10.3389/fpubh.2020.556789] [Medline: 33224912]

18. Abdulaal A, Patel A, Charani E, Denny S, Mughal N, Moore L. Prognostic Modeling of COVID-19 Using Artificial Intelligence in the United Kingdom: Model Development and Validation. J Med Internet Res 2020 Aug 25;22(8):e20259 [FREE Full text] [doi: 10.2196/20259] [Medline: 32735549]
19. Hu R, Gong Y, Guo Y. Federated Learning with Sparsification-Amplified Privacy and Adaptive Optimization. 2021 Presented at: 30th International Joint Conference on Artificial Intelligence (IJCAI-21). 2021; Aug 19-26; Montreal. [doi: 10.24963/ijcai.2021/202]
20. Hu R, Guo Y, Li H, Pei Q, Gong Y. Personalized Federated Learning With Differential Privacy. IEEE Internet Things J 2020 Oct;7(10):9530-9539. [doi: 10.1109/jiot.2020.2991416]
21. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. PLoS Med 2018 Nov 6;15(11):e1002683 [FREE Full text] [doi: 10.1371/journal.pmed.1002683]
22. Sabanayagam C, Xu D, Ting DSW, Nusinovici S, Banu R, Hamzah H, et al. A deep learning algorithm to detect chronic kidney disease from retinal photographs in community-based populations. The Lancet Digital Health 2020 Jun;2(6):e295-e302. [doi: 10.1016/s2589-7500(20)30063-7]
23. Mamoshina P, Ojomoko L, Yanovich Y, Ostrovski A, Botezatu A, Prikhodko P, et al. Converging blockchain and next-generation artificial intelligence technologies to decentralize and accelerate biomedical research and healthcare. Oncotarget 2018 Jan 19;9(5):5665-5690 [FREE Full text] [doi: 10.18632/oncotarget.22345] [Medline: 29464026]
24. Farrar CR, Worden K. Structural Health Monitoring. Hoboken, New Jersey, USA: John Wiley & Sons, Ltd; 2012.
25. Wiens J, Shenoy ES. Machine Learning for Healthcare: On the Verge of a Major Shift in Healthcare Epidemiology. Clin Infect Dis 2018 Jan 06;66(1):149-153 [FREE Full text] [doi: 10.1093/cid/cix731] [Medline: 29020316]
26. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nat Med 2019 Jan;25(1):24-29. [doi: 10.1038/s41591-018-0316-z] [Medline: 30617335]
27. Alhussein M, Muhammad G. Voice Pathology Detection Using Deep Learning on Mobile Healthcare Framework. IEEE Access 2018;6:41034-41041. [doi: 10.1109/ACCESS.2018.2856238]
28. Petracca F, Ciani O, Cucciniello M, Tarricone R. Harnessing Digital Health Technologies During and After the COVID-19 Pandemic: Context Matters. J Med Internet Res 2020 Dec 30;22(12):e21815 [FREE Full text] [doi: 10.2196/21815] [Medline: 33351777]
29. Koonin LM, Hoots B, Tsang CA, Leroy Z, Farris K, Jolly B, et al. Trends in the Use of Telehealth During the Emergence of the COVID-19 Pandemic - United States, January-March 2020. MMWR Morb Mortal Wkly Rep 2020 Oct 30;69(43):1595-1599 [FREE Full text] [doi: 10.15585/mmwr.mm6943a3] [Medline: 33119561]
30. Lonergan PE, Washington Iii SL, Branagan L, Gleason N, Pruthi RS, Carroll PR, et al. Rapid Utilization of Telehealth in a Comprehensive Cancer Center as a Response to COVID-19: Cross-Sectional Analysis. J Med Internet Res 2020 Jul 06;22(7):e19322 [FREE Full text] [doi: 10.2196/19322] [Medline: 32568721]
31. Liu N, Huang R, Baldacchino T, Sud A, Sud K, Khadra M, et al. Telehealth for Noncritical Patients With Chronic Diseases During the COVID-19 Pandemic. J Med Internet Res 2020 Aug 07;22(8):e19493 [FREE Full text] [doi: 10.2196/19493] [Medline: 32721925]
32. Vashist S, Luong J. Wearable technologies for personalized mobile healthcare monitoring and management. In: Tong R, editor. Wearable Technology in Medicine and Health Care. Cambridge, MA: Academic Press; 2018:235-259.
33. Espay AJ, Hausdorff JM, Sánchez-Ferro, Klucken J, Merola A, Bonato P, Movement Disorder Society Task Force on Technology. A roadmap for implementation of patient-centered digital outcome measures in Parkinson's disease obtained using mobile health technologies. Mov Disord 2019;34(5):657-663. [doi: 10.1002/mds.27671] [Medline: 30901495]
34. Kourtis LC, Regele OB, Wright JM, Jones GB. Digital biomarkers for Alzheimer's disease: the mobile/ wearable devices opportunity. NPJ Digit Med 2019;2:1-9 [FREE Full text] [doi: 10.1038/s41746-019-0084-2] [Medline: 31119198]
35. Kumar S, Nilsen W, Pavel M, Srivastava M. Mobile Health: Revolutionizing Healthcare Through Transdisciplinary Research. Computer 2013 Jan;46(1):28-35. [doi: 10.1109/MC.2012.392]
36. Su Z, Figueiredo M, Jo J, Zheng K, Chen Y. Analyzing Description, User Understanding and Expectations of AI in Mobile Health Applications. 2020 Presented at: AMIA 2021 Annual Symposium; Nov 14-18; Virtual p. 1170-1179.
37. Torous J, Larsen ME, Depp C, Cosco TD, Barnett I, Nock MK, et al. Smartphones, Sensors, and Machine Learning to Advance Real-Time Prediction and Interventions for Suicide Prevention: a Review of Current Progress and Next Steps. Curr Psychiatry Rep 2018 Jun 28;20(7):51. [doi: 10.1007/s11920-018-0914-y] [Medline: 29956120]
38. Liu F, Ngai E, Ju X. Understanding mobile health service use: An investigation of routine and emergency use intentions. International Journal of Information Management 2019 Apr;45:107-117. [doi: 10.1016/j.ijinfomgt.2018.09.004]
39. Wu M. Wearable Technology Applications in Healthcare: A Literature Review: OJNI. HIMSS. 2019. URL: https://www.himss.org/resources/wearable-technology-applications-healthcare-literature-review [accessed 2021-12-16]
40. Kim J, Campbell AS, de Ávila BE, Wang J. Wearable biosensors for healthcare monitoring. Nat Biotechnol 2019 Apr;37(4):389-406. [doi: 10.1038/s41587-019-0045-y] [Medline: 30804534]
41. Yang G, Pang G, Pang Z, Gu Y, Mantysalo M, Yang H. Non-Invasive Flexible and Stretchable Wearable Sensors With Nano-Based Enhancement for Chronic Disease Care. IEEE Rev. Biomed. Eng 2019;12:34-71. [doi: 10.1109/rbme.2018.2887301]

42. Ravi D, Wong C, Lo B, Yang G. A Deep Learning Approach to on-Node Sensor Data Analytics for Mobile or Wearable Devices. IEEE J. Biomed. Health Inform 2017 Jan;21(1):56-64. [doi: 10.1109/jbhi.2016.2633287]
43. Yin H, Jha NK. A Health Decision Support System for Disease Diagnosis Based on Wearable Medical Sensors and Machine Learning Ensembles. IEEE Trans. Multi-Scale Comp. Syst 2017 Oct 1;3(4):228-241. [doi: 10.1109/tmscs.2017.2710194]
44. Baig MM, GholamHosseini H, Moqeem AA, Mirza F, Lindén M. A Systematic Review of Wearable Patient Monitoring Systems – Current Challenges and Opportunities for Clinical Adoption. J Med Syst 2017 Jun 19;41(7):115. [doi: 10.1007/s10916-017-0760-1]
45. Friis-Healy EA, Nagy GA, Kollins SH. It Is Time to REACT: Opportunities for Digital Mental Health Apps to Reduce Mental Health Disparities in Racially and Ethnically Minoritized Groups. JMIR Ment Health 2021 Jan 26;8(1):e25456. [doi: 10.2196/25456]
46. Bhatt P, Guo Y, Gong Y. Discounting Lives: Analyzing the causes of Hispanic Adversity raging in the Healthcare sector during Covid-19 using Cumulative Inequality Theory. 2021 Presented at: Proceedings of the Annual Hawaii International Conference on System Sciences. Published online 2021; Jan 5-8; Virtual. [doi: 10.24251/hicss.2021.483]
47. Cai L, Wu C, Meimandi K, Gerber M. Adaptive mobile behavior change intervention using reinforcement learning. 2017 Presented at: International Conference on Companion Technology (ICCT). Published online September; Sep 11-13; Germany. [doi: 10.1109/companion.2017.8287078]
48. Feng S, Duarte MF. Few-shot learning-based human activity recognition. Expert Systems with Applications 2019 Dec;138:112782. [doi: 10.1016/j.eswa.2019.06.070]
49. Wosik J, Fudim M, Cameron B, Gellad ZF, Cho A, Phinney D, et al. Telehealth Transformation: COVID-19 and the rise of Virtual Care. J Am Med Inform Assoc 2020 Apr 20:957-962 [FREE Full text] [doi: 10.1093/jamia/ocaa067] [Medline: 32311034]
50. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019 Jun;6(2):94-98 [FREE Full text] [doi: 10.7861/futurehosp.6-2-94] [Medline: 31363513]
51. Ramkumar PN, Haeberle HS, Bloomfield MR, Schaffer JL, Kamath AF, Patterson BM, et al. Artificial Intelligence and Arthroplasty at a Single Institution: Real-World Applications of Machine Learning to Big Data, Value-Based Care, Mobile Health, and Remote Patient Monitoring. J Arthroplasty 2019 Oct;34(10):2204-2209. [doi: 10.1016/j.arth.2019.06.018] [Medline: 31280916]
52. Laï MC, Brian M, Mamzer M. Perceptions of artificial intelligence in healthcare: findings from a qualitative survey study among actors in France. J Transl Med 2020 Jan 09;18(1):14 [FREE Full text] [doi: 10.1186/s12967-019-02204-y] [Medline: 31918710]
53. Oh S, Kim JH, Choi S, Lee HJ, Hong J, Kwon SH. Physician Confidence in Artificial Intelligence: An Online Mobile Survey. J Med Internet Res 2019 Mar 25;21(3):e12422. [doi: 10.2196/12422]
54. Antoniadou E, Belo D, D'Silva KJ, Wang B, Soboczenski F. Enabling autonomous clinical decision support systems in space through AI-enhanced wearables. In: ICML 2020. Healthcare Systems, Population Health, and the Role of Health-Tech. 2020 Presented at: Proceedings of the International Conference on Machine Learning; 13-18 July; Virtual. [doi: 10.4018/978-1-4666-9432-3.ch010]
55. Chan J, Rea T, Gollakota S, Sunshine JE. Contactless cardiac arrest detection using smart devices. npj Digit. Med 2019 Jun 19;2(1):1-8. [doi: 10.1038/s41746-019-0128-7]
56. Jiang J, Cameron A. IT-Enabled Self-Monitoring for Chronic Disease Self-Management: An Interdisciplinary Review. MISQ 2020 Jan 01;44(1):451-508. [doi: 10.25300/misq/2020/15108]
57. Ju C, Zhao R, Sun J, Wei X, Zhao B, Liu Y, et al. Privacy-preserving technology to help millions of people: Federated prediction model for stroke prevention. 2020 Presented at: International Workshop on Federated Learning for User Privacy and Data Confidentiality at IJCAI 2020; Jan 8, 2021; Virtual.
58. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
59. Tang C, Perez-Pozuelo I, Spathis D, Mascolo C. Exploring contrastive learning in human Activity Recognition for healthcare. 2020 Presented at: Machine Learning for Mobile Health Workshop at NeurIPS 2020; Dec 12; Virtual.
60. Brohman K, Addas S, Dixon J, Pinsonneault A. Cascading Feedback: A Longitudinal Study of a Feedback Ecosystem for Telemonitoring Patients with Chronic Disease. MISQ 2020 Jan 01;44(1):421-450. [doi: 10.25300/misq/2020/15089]
61. Tailor S, Chauhan J, Mascolo C. A First Step Towards On-Device Monitoring of Body Sounds in the Wild. : ACM; 2020 Presented at: Adjunct Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing; Sep 13; Virtual. [doi: 10.1145/3410530.3414440]
62. Van ST, Deschrijver D, Dhaene T. Sensor fusion using backward shortcut connections for sleep apnea detection in multi-modal data. In: arXiv csLG. 2020 Presented at: NeurIPS 2019 Machine Learning for Health Workshop; Dec 13, 2019; Virtual p. 112-125.
63. Bent B, Dunn JP. Personalized Machine Learning Models for Noninvasive Glucose Prediction Using Wearables. In: Machine Learning for Mobile Health Workshop at NeurIPS 2020. 2020 Presented at: Machine Learning for Mobile Health Workshop at NeurIPS 2020; Dec 12; Virtual.

64. Flagg C, Frieder O, MacAvaney S, Motamedi G. Real-time streaming of gait assessment for Parkinson's disease. In: Proceedings of the 14th ACM International Conference on Web Search and Data Mining. ACM; 2021 Presented at: 14th ACM International Conference on Web Search and Data Mining; Mar 8-12; Virtual. [doi: 10.1145/3437963.3441701]
65. Son J, Flatley Brennan P, Zhou S. A Data Analytics Framework for Smart Asthma Management Based on Remote Health Information Systems with Bluetooth-Enabled Personal Inhalers. MISQ 2020 Jan 01;44(1):285-303. [doi: 10.25300/misq/2020/15092]
66. Zhang W, Ram S. A Comprehensive Analysis of Triggers and Risk Factors for Asthma Based on Machine Learning and Large Heterogeneous Data Sources. MISQ 2020 Jan 01;44(1):305-349. [doi: 10.25300/misq/2020/15106]
67. Yamaguchi S, Mino Y, Uddin S. Strategies and future attempts to reduce stigmatization and increase awareness of mental health problems among young people: a narrative review of educational interventions. Psychiatry Clin Neurosci 2011 Aug;65(5):405-415. [doi: 10.1111/j.1440-1819.2011.02239.x] [Medline: 21851449]
68. Lyon EN, Victor LH, Sano A. Bayesian Hierarchical Vector Autoregression Models for Health Label and Mobile Sensor-based Behavioral Feature Prediction. In: NeurIPS 2020 Workshop, Machine Learning for Mobile Health. 2020 Presented at: Proceedings of the NeurIPS Machine Learning for Mobile Health Workshop; Dec 11; Virtual. [doi: 10.1109/embc46164.2021.9630732]
69. Rauber J, Fox E, Gatys L. Modeling patterns of smartphone usage and their relationship to cognitive health. 2019 Presented at: Machine Learning for Health Workshop at NeurIPS 2019; Dec 13; Virtual.
70. Siniver E, Yaniv G. Optimism, pessimism, mood swings and dishonest behavior. Journal of Economic Psychology 2019 Jun;72:54-63. [doi: 10.1016/j.joep.2019.01.007]
71. Calvo R, Dinakar K, Picard R, Maes P. Computing in Mental Health. In: Computing in Mental Health. 2016 Presented at: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems; May 7-12, 2016; San Jose, California, USA. [doi: 10.1145/2851581.2856463]
72. Liu T, Liang P, Muszynski M. Multimodal privacy-preserving mood prediction from mobile data: A preliminary study. In: arXiv csLG. 2020 Presented at: NeurIPS 2020 Workshop on Machine Learning for Mobile Health; Dec 12; Virtual.
73. Moreno-Muñoz P, Romero-Medrano L, Moreno, Herrera-López J, Baca-García E, Artés-Rodríguez A. Passive detection of behavioral shifts for suicide attempt prevention. 2020 Presented at: NeurIPS 2020 Machine Learning for Mobile Health Workshop; Dec 12; Virtual.
74. Huckins JF, daSilva AW, Wang R, Wang W, Hedlund EL, Murphy EI, et al. Fusing Mobile Phone Sensing and Brain Imaging to Assess Depression in College Students. Front Neurosci 2019;13:248 [FREE Full text] [doi: 10.3389/fnins.2019.00248] [Medline: 30949024]
75. Huckins JF, daSilva AW, Wang W, Hedlund E, Rogers C, Nepal SK, et al. Mental Health and Behavior of College Students During the Early Phases of the COVID-19 Pandemic: Longitudinal Smartphone and Ecological Momentary Assessment Study. J Med Internet Res 2020 Jun 17;22(6):e20185 [FREE Full text] [doi: 10.2196/20185] [Medline: 32519963]
76. Chau M, Li TMH, Wong PWC, Xu JJ, Yip PSF, Chen H. Finding People with Emotional Distress in Online Social Media: A Design Combining Machine Learning and Rule-Based Classification. MISQ 2020 Jun 1;44(2):933-955. [doi: 10.25300/misq/2020/14110]
77. Ragav A, Gudur G. Bayesian Active Learning for wearable stress and affect detection. 2020 Presented at: NeurIPS 2020 Machine Learning for Mobile Health Workshop; Dec 12; Virtual.
78. Spathis D, Perez-Pozuelo I, Brage S, Wareham N, Mascolo C. Learning generalizable physiological representations from large-scale wearable data. 2020 Presented at: NeurIPS 2020 Machine Learning for Mobile Health workshop; Dec. 12, 2020; Virtual URL: http://arxiv.org/abs/2011.04601 [doi: 10.1145/3450439.3451863]
79. Menictas M, Tomkins S, Murphy S. Fast Physical Activity Suggestions: Efficient Hyperparameter Learning in Mobile Health. 2020 Presented at: Neurips 2020 Machine Learning in Mobile Health Workshop; Dec. 12, 2020; Virtual URL: http://arxiv.org/abs/2012.11646
80. Imteaj A, Amini M. Distributed sensing using smart end-user devices: Pathway to federated learning for autonomous IoT. Distributed sensing using smart end-user devices: Pathway to federated learning for autonomous IoT. InInternational Conference on Computational Science and Computational Intelligence (CSCI). IEEE; 2019 Presented at: International Conference on Computational Science and Computational Intelligence (CSCI); Dec 5-7; Las Vegas. [doi: 10.1109/csci49370.2019.00218]
81. Kerr J, Carlson J, Godbole S, Cadmus-Bertram L, Bellettiere J, Hartman S. Improving hip-worn accelerometer estimates of sitting using machine learning methods. Medicine & Science in Sports & Exercise 2018;50(7):1518-1524. [doi: 10.1249/mss.0000000000001578]
82. Kheirkhahan M, Nair S, Davoudi A, Rashidi P, Wanigatunga AA, Corbett DB, et al. A smartwatch-based framework for real-time and online assessment and mobility monitoring. Journal of Biomedical Informatics 2019 Jan;89:29-40. [doi: 10.1016/j.jbi.2018.11.003]
83. Luo H, Lee P, Clay I, Jaggi M, De Luca V. Assessment of Fatigue Using Wearable Sensors: A Pilot Study. Digit Biomark 2020;4(Suppl 1):59-72 [FREE Full text] [doi: 10.1159/000512166] [Medline: 33442581]

84. Buvik A, Bergmo TS, Bugge E, Smaabrekke A, Wilsgaard T, Olsen JA. Cost-Effectiveness of Telemedicine in Remote Orthopedic Consultations: Randomized Controlled Trial. J Med Internet Res 2019 Feb 19;21(2):e11330 [FREE Full text] [doi: 10.2196/11330] [Medline: 30777845]
85. Wilmink G, Dupey K, Alkire S, Grote J, Zobel G, Fillit HM, et al. Artificial Intelligence-Powered Digital Health Platform and Wearable Devices Improve Outcomes for Older Adults in Assisted Living Communities: Pilot Intervention Study. JMIR Aging 2020 Sep 10;3(2):e19554 [FREE Full text] [doi: 10.2196/19554] [Medline: 32723711]
86. Zhu H, Samtani S, Brown R, Chen H. A Deep Learning Approach for Recognizing Activity of Daily Living (ADL) for Senior Care: Exploiting Interaction Dependency and Temporal Patterns. MISQ 2021 Jun 1;45(2):859-896. [doi: 10.25300/misq/2021/15574]
87. Zhu H, Samtani S, Chen H, Nunamaker JF. Human Identification for Activities of Daily Living: A Deep Transfer Learning Approach. Journal of Management Information Systems 2020 Jun 16;37(2):457-483. [doi: 10.1080/07421222.2020.1759961]
88. Gudur GK, Perepu SK. Federated Learning with heterogeneous labels and models for mobile activity monitoring. 2020 Presented at: Machine Learning for Mobile Health Workshop at NeurIPS 2020; 2020/12/12; Virtual URL: http://arxiv.org/abs/2012.02539
89. Alrumayh A, Tan C. Supporting home quarantine with smart speakers. In: Proceedings of Deep Learning for Wellbeing Applications Leveraging Mobile Devices and Edge Computing. ACM; 2020 Presented at: Deep Learning for Wellbeing Applications Leveraging Mobile Devices and Edge Computing; Jun 19; Virtual. [doi: 10.1145/3396868.3400897]
90. Tang Q, Ponnada A, Intille S. Towards Personal Hand Hygiene Detection in Free-living Using Wearable Devices. 2020 Presented at: Machine Learning for Mobile Health Workshop at Conference on Neural Information Processing Systems (NeurIPS) 2020; Dec 12; Virtual.
91. Feng J, Rong C, Sun F, Guo D, Li Y. PMF: A privacy-preserving human mobility prediction framework via federated learning. In: Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 2020 Mar 18 Presented at: Proc ACM Interact Mob Wearable Ubiquitous Technol. 2020;4(1); March; Virtual p. 1-21. [doi: 10.1145/3381006]
92. Gomis-Pastor M, Mirabet S, Roig E, Lopez L, Brossa V, Galvez-Tugas E, et al. Interdisciplinary Mobile Health Model to Improve Clinical Care After Heart Transplantation: Implementation Strategy Study. JMIR Cardio 2020 Nov 24;4(1):e19065 [FREE Full text] [doi: 10.2196/19065] [Medline: 33231557]
93. Amoedo A, Martnez-Costa MDP, Moreno E. An analysis of the communication strategies of Spanish commercial music networks on the web: http://los40.com, http://los40principales.com, http://cadena100.es, http://europafm.es and http://kissfm.es. radio journal: international studies in 2009 Feb 01;6(1):5-20 [FREE Full text] [doi: 10.1386/rajo.6.1.5_4]
94. Shailaja K, Seetharamulu B, Jabbar M. Machine learning in healthcare: A review. 2018 Presented at: In2018 Second international conference on electronics, communication and aerospace technology (ICECA) (pp.). IEEE; Mar 29; Coimbatore, India p. 910-914. [doi: 10.1109/iceca.2018.8474918]
95. Chen Y, Qin X, Wang J, Yu C, Gao W. FedHealth: A Federated Transfer Learning Framework for Wearable Healthcare. IEEE Intell. Syst 2020 Jul 1;35(4):83-93. [doi: 10.1109/mis.2020.2988604]
96. Konečný J, McMahan B, Ramage D. Federated optimization: Distributed optimization beyond the datacenter. arXiv preprint arXiv.03575 Nov 2015:11. [doi: 10.48550/arXiv.1511.03575]
97. Gunning D, Aha D. DARPA's Explainable Artificial Intelligence (XAI) Program. AIMag 2019 Jun 24;40(2):44-58. [doi: 10.1609/aimag.v40i2.2850]
98. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]
99. Adadi A, Berrada M. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). IEEE Access 2018;6:52138-52160. [doi: 10.1109/access.2018.2870052]
100. Gordon L, Grantcharov T, Rudzicz F. Explainable Artificial Intelligence for Safe Intraoperative Decision Support. JAMA Surg 2019 Nov 01;154(11):1064. [doi: 10.1001/jamasurg.2019.2821]
101. Payrovnaziri S, Chen Z, Rengifo-Moreno P, Miller T, Bian J, Chen JH, et al. Explainable artificial intelligence models using real-world electronic health record data: a systematic scoping review. J Am Med Inform Assoc 2020 Jul 01;27(7):1173-1185 [FREE Full text] [doi: 10.1093/jamia/ocaa053] [Medline: 32417928]

Abbreviations

AAAI: Association for the Advancement of Artificial Intelligence
ACM: Association for Computing Machinery
AI: artificial intelligence
AIM: artificial intelligence–powered mobile health
DARPA: Defense Advanced Research Projects Agency
DL: deep learning
FL: federated learning
IOT: internet of things
IS: information systems

ISR: Information Systems Research
IT: information technology
mHealth: mobile health
MISQ: Management Information Systems Quarterly
ML: machine learning
NeurIPS: Neural Information Processing Systems
NIPS: Neural Information Processing Systems
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RQ: research question
XAI: explainable AI

Edited by L Buis; submitted 18.11.21; peer-reviewed by F Velayati, D Madhusudhan, H Ayatollahi; comments to author 28.11.21; revised version received 23.01.22; accepted 08.04.22; published 09.06.22.

Please cite as:
Bhatt P, Liu J, Gong Y, Wang J, Guo Y
Emerging Artificial Intelligence–Empowered mHealth: Scoping Review
JMIR Mhealth Uhealth 2022;10(6):e35053
URL: https://mhealth.jmir.org/2022/6/e35053
doi:10.2196/35053
PMID:35679107

Review

Quality Evaluation of Free-living Validation Studies for the Assessment of 24-Hour Physical Behavior in Adults via Wearables: Systematic Review

Marco Giurgiu[1,2], PhD; Irina Timm[1], MSc; Marlissa Becker[3], MSc; Steffen Schmidt[1], PhD; Kathrin Wunsch[1], PhD; Rebecca Nissen[1], BSc; Denis Davidovski[1], BSc; Johannes B J Bussmann[4], PhD; Claudio R Nigg[5], Prof Dr; Markus Reichert[1,2,6], PhD; Ulrich W Ebner-Priemer[1,2], Prof Dr; Alexander Woll[1], Prof Dr; Birte von Haaren-Mack[7], PhD

[1]Department of Sports and Sports Science, Karlsruhe Institute of Technology, Karlsruhe, Germany

[2]Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

[3]Unit Physiotherapy, Department of Orthopedics, Erasmus MC, University Medical Center Rotterdam, Rotterdam, Netherlands

[4]Department of Rehabilitation Medicine, Erasmus MC, University Medical Center Rotterdam, Rotterdam, Netherlands

[5]Health Science Department, Institute of Sport Science, University of Bern, Bern, Switzerland

[6]Department of eHealth and Sports Analytics, Faculty of Sport Science, Ruhr-University Bochum, Bochum, Germany

[7]Department of Health and Social Psychology, Institute of Psychology, German Sport University, Cologne, Germany

Corresponding Author:
Marco Giurgiu, PhD
Department of Sports and Sports Science, Karlsruhe Institute of Technology
Hertzstr. 16
Karlsruhe, 76187
Germany
Phone: 49 017620763557
Email: marco.giurgiu@kit.edu

Abstract

Background: Wearable technology is a leading fitness trend in the growing commercial industry and an established method for collecting 24-hour physical behavior data in research studies. High-quality free-living validation studies are required to enable both researchers and consumers to make guided decisions on which study to rely on and which device to use. However, reviews focusing on the quality of free-living validation studies in adults are lacking.

Objective: This study aimed to raise researchers' and consumers' attention to the quality of published validation protocols while aiming to identify and compare specific consistencies or inconsistencies between protocols. We aimed to provide a comprehensive and historical overview of which wearable devices have been validated for which purpose and whether they show promise for use in further studies.

Methods: Peer-reviewed validation studies from electronic databases, as well as backward and forward citation searches (1970 to July 2021), with the following, required indicators were included: protocol must include real-life conditions, outcome must belong to one dimension of the 24-hour physical behavior construct (intensity, posture or activity type, and biological state), the protocol must include a criterion measure, and study results must be published in English-language journals. The risk of bias was evaluated using the Quality Assessment of Diagnostic Accuracy Studies-2 tool with 9 questions separated into 4 domains (patient selection or study design, index measure, criterion measure, and flow and time).

Results: Of the 13,285 unique search results, 222 (1.67%) articles were included. Most studies (153/237, 64.6%) validated an intensity measure outcome such as energy expenditure. However, only 19.8% (47/237) validated biological state and 15.6% (37/237) validated posture or activity-type outcomes. Across all studies, 163 different wearables were identified. Of these, 58.9% (96/163) were validated only once. ActiGraph GT3X/GT3X+ (36/163, 22.1%), Fitbit Flex (20/163, 12.3%), and ActivPAL (12/163, 7.4%) were used most often in the included studies. The percentage of participants meeting the quality criteria ranged from 38.8% (92/237) to 92.4% (219/237). On the basis of our classification tree to evaluate the overall study quality, 4.6% (11/237) of studies were classified as *low risk*. Furthermore, 16% (38/237) of studies were classified as having *some concerns*, and 72.9% (173/237) of studies were classified as *high risk*.

Conclusions: Overall, free-living validation studies of wearables are characterized by low methodological quality, large variability in design, and focus on intensity. Future research should strongly aim at biological state and posture or activity outcomes and strive for standardized protocols embedded in a validation framework. Standardized protocols for free-living validation embedded in a framework are urgently needed to inform and guide stakeholders (eg, manufacturers, scientists, and consumers) in selecting wearables for self-tracking purposes, applying wearables in health studies, and fostering innovation to achieve improved validity.

(JMIR Mhealth Uhealth 2022;10(6):e36377) doi:10.2196/36377

KEYWORDS

wearables; validation; sedentary behavior; physical activity; sleep

Introduction

24-Hour Physical Behavior

Although physical activity (PA) has been commonly assessed using self-report measures in the past decades, device-based measurement of PA research has been on the rise for several years. This method has also developed since its implementation and moved forward from assessing single parameters such as steps and counts, over the assessment of sedentary behavior (SB) and PA in parallel, to an integrated perspective of different movement and nonmovement patterns—the so-called 24-hour activity cycle (24-HAC) [1,2]. Studies have shown that these different parameters independently contribute to health. Hence, the current World Health Organization guidelines [3] encourage adults and older adults to increase the time spent on PA while simultaneously limiting the amount of sedentary time. Positive implications can be expected for overall physical and mental health throughout the life span, including having a healthy sleep pattern to this recommendation.

This apparent shift from investigating a single behavior such as PA to a multi-perspective focus on 24-hour physical behavior (ie, including sleep, SB, and PA) has also been theoretically addressed. Rosenberger et al [1] introduced the 24-HAC model as a new paradigm for PA, and Trembley et al [2] provided a conceptual model of movement-based terminology around the 24-hour cycle. This approach was further extended by the Prospective Physical Activity, Sitting, and Sleep consortium, which suggested a subdivision of the 24-hour physical behavior construct into 3 behaviors by applying different dimensions [4], meaning that each behavior covers aspects of biological (ie, sleep or awake), postural (eg, lying, sitting, and upright), and intensity (eg, light, moderate, and vigorous) dimensions. Therefore, the differentiation among PA, SB, and sleep in terms of the 24-HAC model requires valid and simultaneous assessments of all 3 dimensions (ie, biological state, posture or activity type, and intensity; Multimedia Appendix 1, Table S1 [5-226]) under real-life conditions. Rigorous validation studies performed in a free-living environment are necessary to accurately predict the performance of a device and algorithm under real-life conditions.

Wearable Technology and Validation

As technical opportunities have evolved rapidly during the past decades, wearables (ie, body-worn devices such as accelerometers, smartwatches, pedometers, or fitness trackers) have become a leading fitness trend, with an estimated US $95 billion industry [227], which is still growing. Moreover, a considerable number of research studies integrated device-based methods to capture physical behavior data, and first discussions have already come up on whether it is *prime time* for wearables with scientific validation to be a global physical behavior surveillance methodology [228,229]. However, the application of wearables in studies that assess health-related questions presents several methodological and practical challenges. For example, strategies for data processing, monitoring protocols, assessment limitations (eg, muscle-strengthening exercises), and quality criteria such as validity need to be taken into account [230].

An important test quality criterion is the concept of validity, which represents a fundamental criterion for evaluating the quality of an instrument, referring to the degree to which it truly measures the construct it targets [231]. Regarding the 24-hour physical behavior cycle, researchers are commonly interested in criterion-referenced validity as their assessed outcome parameters are highly objective [232]. Although (or because of) the number of validation studies has increased over the past years, there is high heterogeneity across published protocols and used measurement methods, which severely limits valid device comparisons [233]. Thus, suggestions for standardized validation procedures have received increasing attention in the scientific community [233-235].

Standardized Protocols and Validation Framework

There have been several attempts in this direction, as collaborations such as the INTERLIVE network have already started developing standardized protocols to validate consumer wearables for steps [233] and heart rate [236]. Furthermore, Keadle et al [234] introduced a stage process framework of validity to facilitate the development and validation of processing methods for assessing physical behavior using wearables. This framework contains 5 validation phases with increasing levels, starting from device manufacturing and culminating with application in health studies. Validation studies should be implemented following mechanical testing (phase 0) and calibration testing (phase 1). Here, a fixed and semistructured evaluation under laboratory conditions (phase 2) should be applied, followed by an evaluation under real-life conditions (phase 3) [234]. The validation of devices should pass through all these stages before the respective device can be used in health research studies (phase 4). As there is a nonnegligible difference in error rates between laboratory and real-life conditions [233], a wide array of activities of daily living should be captured and compared under real-life conditions. It also needs to be taken into account that participants are instructed to perform specific activities under laboratory

conditions, which may result in unnaturally performed activities (eg, the Hawthorne effect) [237]. Hence, the quantification of measurement error is essential to be defined in an unconstrained free-living environment, and wearables' outcomes should be compared with a reference measure such as video recordings or the doubly labeled water method, depending on the outcome parameter of choice. Overall, the aim should be the realization of standardized validation protocols to be embedded in a framework [233,234], which may have positive implications for all stakeholders such as manufacturers, scientists, and consumers. The results of validation studies are helpful to disabuse consumers and can assist researchers in study design when selecting an appropriate wearable device for the respective question or questions to be answered [238,239].

Objectives

Although validated devices are a prerequisite for proper research and validation frameworks have been proposed, to the best of our knowledge, no previous review has systematically evaluated the characteristics and quality of free-living validation studies. This review focuses on the following purposes. First, as our main purpose, we aimed to raise researchers' and consumers' attention to the quality of published validation protocols while aiming to identify and compare specific consistencies and inconsistencies between validation protocols. To evaluate the quality of the studies, we followed core principles, recommendations, and expert statements [232-234,240] with published quality criteria (eg, study duration, number of included participants, selection of criterion measures, and data synchronization). Second, we aimed to provide a comprehensive and historical overview of which wearable devices have been validated for which purpose and whether they show promise for use in further studies.

Methods

This study followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) reporting guidelines (Multimedia Appendix 1, Table S2 [5-226]) [241].

Search Strategy and Study Selection

Three separate search strings were combined: terms for validation analyses, 24-hour physical behavior constructs, and wearables. An a priori pilot search was conducted to optimize the final term (Multimedia Appendix 1, Table S3 [5-226]). Publications from 1970 to December 2020 were searched using the following databases: EBSCOhost, IEEE Xplore, PubMed, Scopus, and Web of Science. We reran the search in July 2021 to check for updates and checked the reference lists of included studies for publications that met the inclusion criteria.

All articles were imported to the Citavi library (Citavi 6.8; Swiss Academic Software GmbH). After removing all duplicates, the study selection process included 3 screening phases for eligibility. In the first phase, 2 reviewers (MG and RN) independently screened the titles of the publications. Articles were excluded only if both reviewers categorized them as not eligible for review purposes. In the second phase, 2 reviewers (MG and RN) independently screened and reviewed the abstracts of the publications. Discrepancies in screening were resolved by consulting with a third reviewer (BvHM). Finally, in the third phase, the full texts of the remaining articles were assessed for eligibility by 7 members of the author's team (MG, RN, DD, KS, SS, IT, and BvHM). Each article was independently screened by at least two reviewers. Discrepancies in screening were resolved through discussion until a consensus was reached. The reviewers were not blinded to the author or journal information.

Inclusion and Exclusion Criteria

On the basis of the population, intervention, comparison, and outcome principle [242], we included peer-reviewed English-language publications that met the criteria described in the following sections.

Population

Participants were adults and older adults aged ≥18 years, regardless of health conditions. Studies that specifically targeted adults and populations of older adults (aged ≥18 years) were excluded.

Intervention

Any wearable validation study in which at least one part of the study was conducted under free-living (naturalistic or real-life) conditions (eg, at participants' homes or schools and without instructions on when to start or stop a particular activity) was included. Studies in which the protocol was conducted under laboratory conditions were excluded.

Control or Comparison

We included only studies where a criterion measure was described (eg, observation or wearable devices) and excluded all studies where no criterion measure was described (eg, comparison between 2 devices without indicating a criterion measure).

Outcomes

Studies were included in which the wearable outcome or outcomes could be classified into at least one dimension of the 24-hour physical behavior construct (ie, biological state, posture or activity type, or intensity [4]; Multimedia Appendix 1, Table S1 [5-226]). We excluded studies in which the outcome did not distinguish among sleep-wake states; intensities; or posture or activity types such as sleep quality, gait analyses, or heart rate parameters.

Data Extraction and Synthesis

Data extraction and synthesis were conducted independently by 2 authors (MG, RN, DD, KW, SS, or IT). Discrepancies were discussed until a consensus was reached. The following study details were extracted: author, year, location, population information (sample size, mean age of participants, percentage of females, and ethnicity), measurement period, validated wearable device (wearing position, software, epoch length, and cutoff point or algorithm), dimensions of the 24-hour physical behavior construct (biological state, intensity, and posture or activity type), validated outcome, criterion measure, statistical analyses for validation purposes, conclusion, and funding information. Given the wide range of study protocols in terms of varying conditions (eg, wear location, measurement duration,

sample size, statistical analyses, or criterion measure), we conducted a narrative synthesis based on the reported results or conclusions. The data synthesis focused on the purpose (ie, whether the included wearables showed promise for use in further studies). In particular, we classified the studies as moderate to strong validity, mixed results, and poor or weak validity.

Quality Assessment

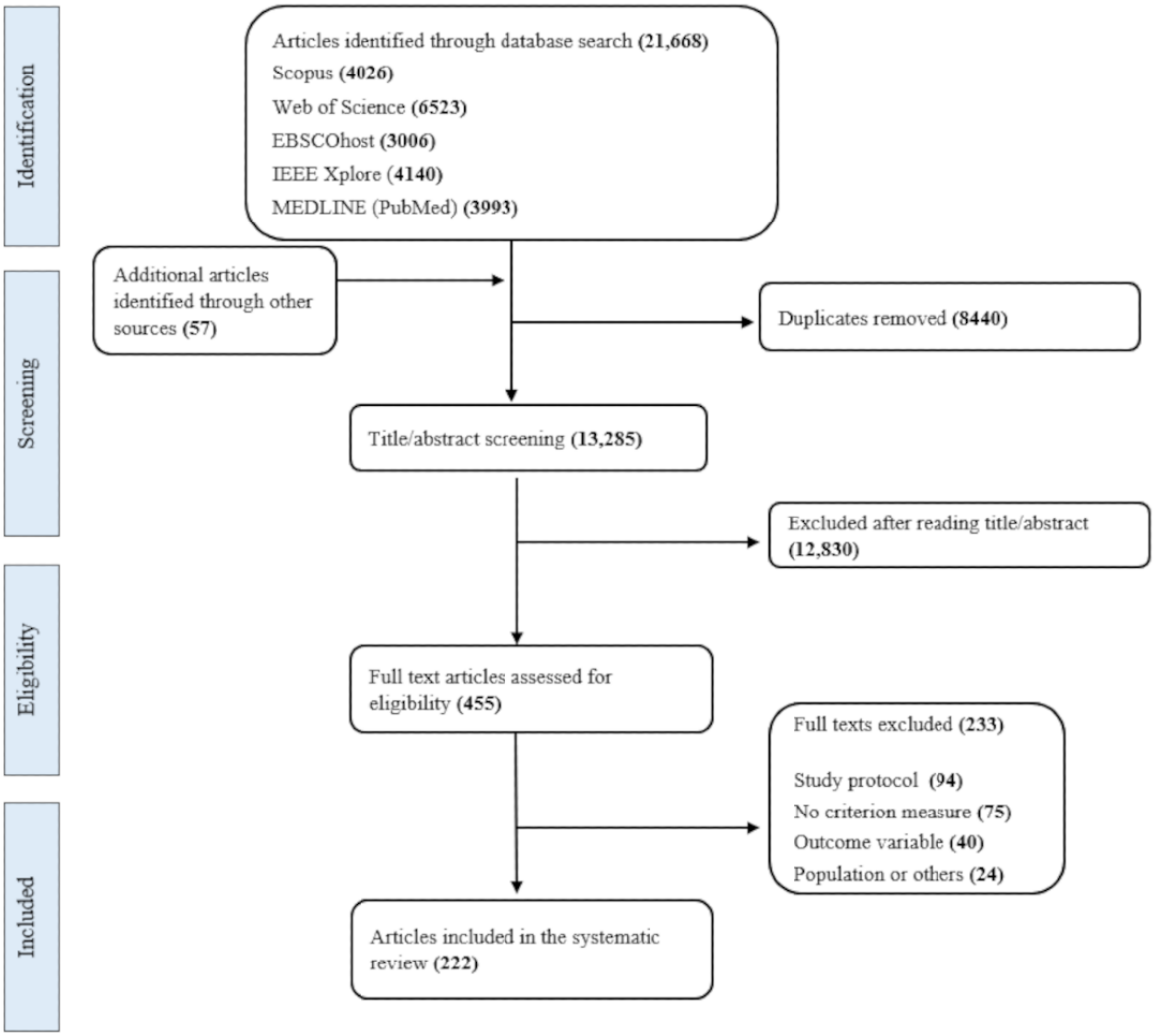
The risk of bias for each article was evaluated using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool [240]. The tool comprises 4 different domains (ie, patient selection, index measure, criterion measure, and flow and timing). Following the QUADAS-2 guidelines, we selected a set of signaling questions for each domain and added questions modified from the QUADAS-2 background document based on core principles, recommendations, and expert statements for validation studies [232-234,240]. The risk of bias assessment was independently conducted by at least two authors. Discrepancies were discussed until a consensus was reached. The study quality was evaluated at the domain level; that is, if all signaling questions for a domain were answered *yes*, then the risk of bias was deemed to be *low*. If any signaling question was answered *no*, then the risk of bias was deemed to be *high*. The *unclear* category was only used when insufficient data were reported for evaluation. On the basis of domain-level ratings, we created a decision tree to evaluate the overall study quality as *low risk*, *some concerns*, or *high risk* (Multimedia Appendix 1, Figure S1 [5-226]).

Results

Overview

The search resulted in 13,285 unique records, with 222 (1.67%) publications being included [5-226] (Figure 1). Most studies (208/222, 93.7%) validated an outcome from one dimension, whereas few (14/222, 6.3%) studies validated outcomes from 2 different dimensions (ie, intensity and posture or activity type or intensity and biological state) at the same time during a study protocol. Only 0.5% (1/222) of studies included outcomes from all the 3 dimensions. In particular, of all the 237 identified outcomes, 153 (64.6%) were classified into the intensity dimension, 38 (16%) were classified into the posture or activity type dimension, and 47 (21.2%) were classified into the biological state dimension.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart illustrating the literature search and screening process.



Participant and Study Characteristics

Of the included studies, 84.2% (187/222) were published within the past decade (in or after 2011), indicating the increasing use of wearable technologies for physical behavior measurement (Table 1); 93.2% (207/222) were conducted in wealthier high-income countries in North America, Europe, or Australia and Oceania. The number of participants ranged between 1 and 3752, although most studies (113/222, 50.9%) recruited between 20 and 50 participants. The mean age of the participant samples ranged between young (18.0, SD 0.6 years) and older adults (86.4, SD 6.0 years). In most studies, the mean age of the sample was between 18 and 64 years, and the proportion of female participants ranged from 26% to 74% (144/222, 64.8%). Healthy participants were recruited in 79.7% (177/222) of all studies, whereas 20.7% (46/222) of all studies included participants with different physical and mental health restrictions such as cardiometabolic diseases or chronic heart failure in 2.7% (6/222), chronic obstructive pulmonary disease in 2.7% (6/222), stroke in 2.3% (5/222), insomnia in 2.3% (5/222), or intellectual and visual disabilities in 0.9% (2/222). Information about the participants' ethnicity was reported in 14.9% (33/222) of all studies. The conceptualization of study protocols regarding measurement duration varied between approximately 30 minutes and up to several weeks. The study duration of ≤1 day was predominantly for studies that focused on posture or activity-type outcomes (ie, in 18/36, 50% of the included studies). Most studies (205/222, 92.3%) conducted statistical analyses at the person or study level (eg, correlations, 2-tailed *t* tests, and repeated-measures ANOVA). Some studies (21/222, 9.5%) conducted both person or study-level analyses and epoch-by-epoch comparisons (eg, sensitivity and specificity). Approximately 10.4% (23/222) of studies reported that the manufacturer was involved in the study funding or provided devices for validation purposes. No funding information was reported by 16.2% (36/222) of the studies, whereas the remaining studies (164/222, 73.9%) indicated that funding was independent of manufacturer companies. Detailed data extraction is reported as a supplement (Multimedia Appendix 1, Table S4 [5-226]).

Table 1. Summary of data extraction: participant and study characteristics (N=237).

Category	Total, n (%)	Biological state (n=47), n (%)	Posture or activity type (n=37), n (%)	Intensity (n=153), n (%)
Publication year				
Before or in 1999	7 (3)	1 (2.1)	3 (8.1)	3 (2)
2000-2010	28 (11.8)	4 (8.5)	3 (8.1)	21 (13.7)
After or in 2011	187 (78.9)	42 (89.4)	31 (83.8)	129 (84.3)
Study location[a]				
Africa	1 (0.4)	—[b]	—	1 (0.7)
Asia	17 (7.2)	5 (10.6)	2 (5.4)	10 (6.5)
Europe	95 (40.1)	14 (29.8)	21 (57)	69 (45.1)
North America	92 (38.8)	21 (44.7)	10 (27)	65 (42.5)
Australia or Oceania	16 (6.8)	7 (14.9)	4 (10.8)	7 (4.6)
Number of participants				
≤19	71 (30)	12 (25.5)	22 (59.5)	40 (26.1)
20-50	113 (47.7)	23 (48.9)	13 (35.1)	86 (56.2)
≥51	38 (16)	12 (25.5)	2 (5.4)	27 (17.6)
Age (years; mean age)[c]				
18-64	174 (73.4)	36 (76.6)	28 (75.7)	122 (79.7)
≥65	44 (18.6)	10 (21.3)	7 (18.9)	31 (20.3)
Sex (female; %)[d]				
0-25	33 (13.9)	6 (12.8)	6 (16.2)	22 (14.4)
26-74	144 (60.8)	36 (76.6)	22 (59.5)	99 (64.7)
75-100	35 (14.8)	5 (10.6)	8 (21.6)	23 (15)
Measurement duration (days)[e]				
≤1	69 (29.1)	17 (36.2)	19 (51.4)	39 (25.5)
2-6	50 (21.1)	15 (31.9)	7 (18.9)	33 (21.6)
≥7	101 (42.6)	15 (31.9)	9 (24.3)	80 (52.3)
Criterion measure				
Doubly labeled water	42 (17.7)	—	—	42 (27.5)
Heart telemetry	—	—	—	—
Indirect calorimetry	4 (1.7)	—	—	4 (2.6)
Observation (direct)	7 (3)	—	4 (10.8)	3 (2)
Observation (images)	2 (0.8)	—	2 (5.4)	—
Observation (video)	14 (5.9)	—	11 (29.7)	5 (3.3)
Polysomnography	24 (10.1)	24 (51.1)	—	—
Questionnaire or diary	16 (6.8)	6 (12.8)	4 (10.8)	6 (3.9)
Wearable	113 (47.7)	14 (29.8)	16 (43.2)	94 (61.4)
EEG[f] or Zmachine	4 (1.7)	3 (6.4)	—	—
Statistical analyses				
Epoch-by-epoch	33 (13.9)	13 (27.7)	15 (40.5)	8 (5.2)
Person or study level	208 (87.8)	44 (93.6)	28 (75.7)	150 (98)

[a]One study did not report the study location.

[b]Not available.

[c]A total of 5 studies were not included in the summary statistics because of the lack of age information. One study was counted twice as it included 2 different age groups.

[d]A total of 10 studies was not included in the summary statistics because of the lack of sex information.

[e]A total of 2 studies was not included in the summary statistics because of the lack of measurement duration information.

[f]EEG: electroencephalogram.

Wearables

We identified 163 different wearables from 82 companies, of which 61 (37.4%) were classified as research-grade devices and 102 (62.6%) were classified as commercial-grade devices. The types of wearables varied across uniaxial, biaxial, or triaxial accelerometers and pedometers. On the basis of our narrative data synthesis, we ranked 26.4% (122/463 of results or conclusions as moderate to strong validity, 48.8% (226/463) as mixed validity, and 25% (115/463) as poor or weak validity (Multimedia Appendix, Table S5 [5-226]). In relation to other types of wearables, triaxial accelerometers were used in 70.5% (167/237) of all included studies. Detailed technical information for each wearable device is available as a supplement (Multimedia Appendix, Table S6 [5-226]). Of the 163 different wearables, 96 (58.9%) were validated only once. ActiGraph GT3X/GT3X+ (36/163, 22.1%), Fitbit Flex (20/163, 12%), and ActivPAL (12/163, 7.4%) were used most often in the included validation studies. Of all the 222 reviewed articles, 78 (35.1%) studies included different types of wearables, and 26 (11.7%) studies included different wearing positions to enable inter- and intradevice comparisons (Table 2). The variation of different sensor models within a study protocol ranged from 1 to 12 different wearables. In particular, 64.9% (144/222) of all studies included one model of wearable, 20.3% (45/222) included 2 different models of wearables, and 14.9% (33/222) included ≥3 different models of wearables. We identified 11 different validated outcomes. Of all reported outcomes, 51.3% (164/320) represented continuous parameters such as steps, energy expenditure, or counts, whereas 48.8% (156/320) represented categorical outcomes such as sleep time, time spent in light PA, or time spent in moderate to vigorous PA. Approximately 18% (40/222) of studies validated 2 different outcomes such as steps and energy expenditure during a study protocol. More than half of the studies (125/237, 52.7%) validated one type of wearable device in a single wearing position. We identified 13 different wearing positions. The wrist and hip or waist positions were used most often for validation purposes. In 50% (111/222) of all studies, the authors provided information about the software application used for data preprocessing.

Table 2. Summary of data extraction: wearables (N=237).

Category	Total, n (%)	Biological state (n=47), n (%)	Posture or activity type (n=37), n (%)	Intensity (n=153), n (%)
Type				
Uniaxial accelerometer	44 (10.8)	13 (14.4)	3 (6.4)	29 (9.2)
Biaxial accelerometer	28 (6.9)	1 (1.1)	2 (4.3)	26 (8.3)
Triaxial accelerometer	286 (70.1)	72 (80)	40 (85.1)	212 (67.5)
Pedometer	30 (7.4)	2 (2.2)	—[a]	30 (9.6)
Unclear	20 (4.9)	2 (2.2)	2 (4.3)	17 (5.4)
Outcome				
Sleep time	42 (13.1)	42 (89.4)	—	—
Sleep-wake metrics	5 (1.6)	5 (10.6)	—	—
Different postures or types	29 (9.1)	—	30 (73.2)	—
Sit-to-stand transitions	5 (1.6)	—	5 (12.2)	—
Time in sedentary behavior	32 (10)	—	6 (14.6)	26 (11.2)
Time in light physical activity	14 (4.4)	—	—	14 (6)
Time in moderate to vigorous physical activity	33 (10.3)	—	—	33 (14.2)
Time in physical activity	6 (1.9)	—	—	6 (2.6)
Energy expenditure	72 (22.5)	—	—	72 (30.9)
Steps	75 (23.4)	—	—	75 (32.2)
Counts	7 (2.2)	—	—	7 (3)
Wear position[b]				
Ankle	11 (2.4)	1 (1)	1 (1.6)	9 (2.6)
Backpack, pocket, and bra	20 (4.3)	—	2 (3.1)	19 (5.6)
Chest	15 (3.3)	1 (1)	3 (4.7)	11 (3.2)
Foot	3 (0.7)	1 (1)	—	2 (0.6)
Hip and waist	148 (32.2)	7 (7.1)	23 (35.9)	124 (36.4)
Leg	3 (0.7)	—	3 (4.7)	—
Lower back	12 (2.6)	1 (1)	3 (4.7)	10 (2.9)
Neck	3 (0.7)	—	—	3 (0.9)
Thigh	28 (6.1)	2 (2)	15 (23.4)	13 (3.8)
Torso	1 (0.2)	—	—	1 (0.3)
Trunk	3 (0.7)	—	3 (4.7)	—
Upper arm	12 (2.6)	1 (1)	1 (1.6)	10 (2.9)
Wrist	201 (43.7)	83 (84.7)	10 (15.6)	139 (40.8)

[a]Not available.

[b]One study did not report any information about the sensor wearing position and one study did not specify the information about the sensor wearing position. If studies included multiple devices or different wearing positions, we counted each device and wearing position separately.

Study Quality

In total, we included 9 signaling questions as quality criteria to evaluate the risk of bias. The percentage of studies that met the criteria ranged from 38.7% (92/238) to 92% (219/238; Table 3). On average, 5.2 (SD 1.41) of 9 questions were answered with *yes* (ie, meeting the criteria). Studies validating a biological state, intensity, or posture or activity type outcome met on average 4.6 (SD 1.23), 5.5 (SD 1.36), and 4.9 (SD 1.51) out of 9 questions with *yes* (ie, no risk of bias), respectively. We evaluated whether the reference standard was the appropriate gold standard as a central criterion for evaluating overall study quality. In 38.4% (91/237) of all studies, the reference standard was equivalent to the suggested criterion measures [234]. Wearables were the most frequently selected reference criterion in 50.5% (112/222) of the studies. On the basis of our

classification tree to evaluate the overall study quality (Multimedia Appendix 1, Figure S1 and Table S7 [5-226]), 4.6% (11/237) of studies were classified as *low risk*. Furthermore, 16% (38/237) of studies were classified as having *some concerns*, and 72.9% (173/237) of studies were classified as *high risk*. To provide an overview of the study quality, Figure 2 illustrates the overall study quality on a study level, separated by each dimension of the 24-hour physical behavior construct.

Table 3. Criteria for the risk of bias assessment and the percentage of studies meeting these criteria (N=237).

Criteria items	Studies meeting criterion, n (%)			
	Total	Biological state (n=47)	Posture or activity type (n=37)	Intensity (n=153)
Domain 1: patient selection or study design				
Was the study conducted in different free-living settings (eg, work or home)?	174 (91.6)[a]	N/A[b]	28 (75.7)	146 (95.4)
Did the study take place for at least 2 days?	156 (65.8)	28 (59.6)	16 (43.2)	112 (73.2)
Did the study provide any information about the inclusion and exclusion criteria of the recruiting process?	165 (69.6)	34 (72.3)	25 (67.6)	106 (69.3)
Did the study include a sample of at least 20 participants?	163 (68.8)	36 (76.6)	15 (40.5)	112 (73.2)
Domain 2 : index measure				
Was the algorithm of the validated outcome reported (ie, formula) or at least further information cited?	97 (40.9)	23 (48.9)	18 (48.6)	56 (36.6)
Did the participants wear the wearable for at least 8 hours per day?	107 (56.3)[a]	N/A	15 (40.5)	92 (60.1)
Domain 3: criterion measure				
Is the selected reference the gold standard?	91 (38.4)	23 (48.9)	14 (37.8)	54 (35.3)
Domain 4: flow and timing				
Did the authors provide any information about data synchronization?	75 (41.4)[c]	24 (51.1)	22 (59.5)	29 (19)
Were all participants included in the analyses or were any exclusion reasons provided?	218 (92.4)[d]	45 (95.7)	31 (83.8)	142 (92.8)

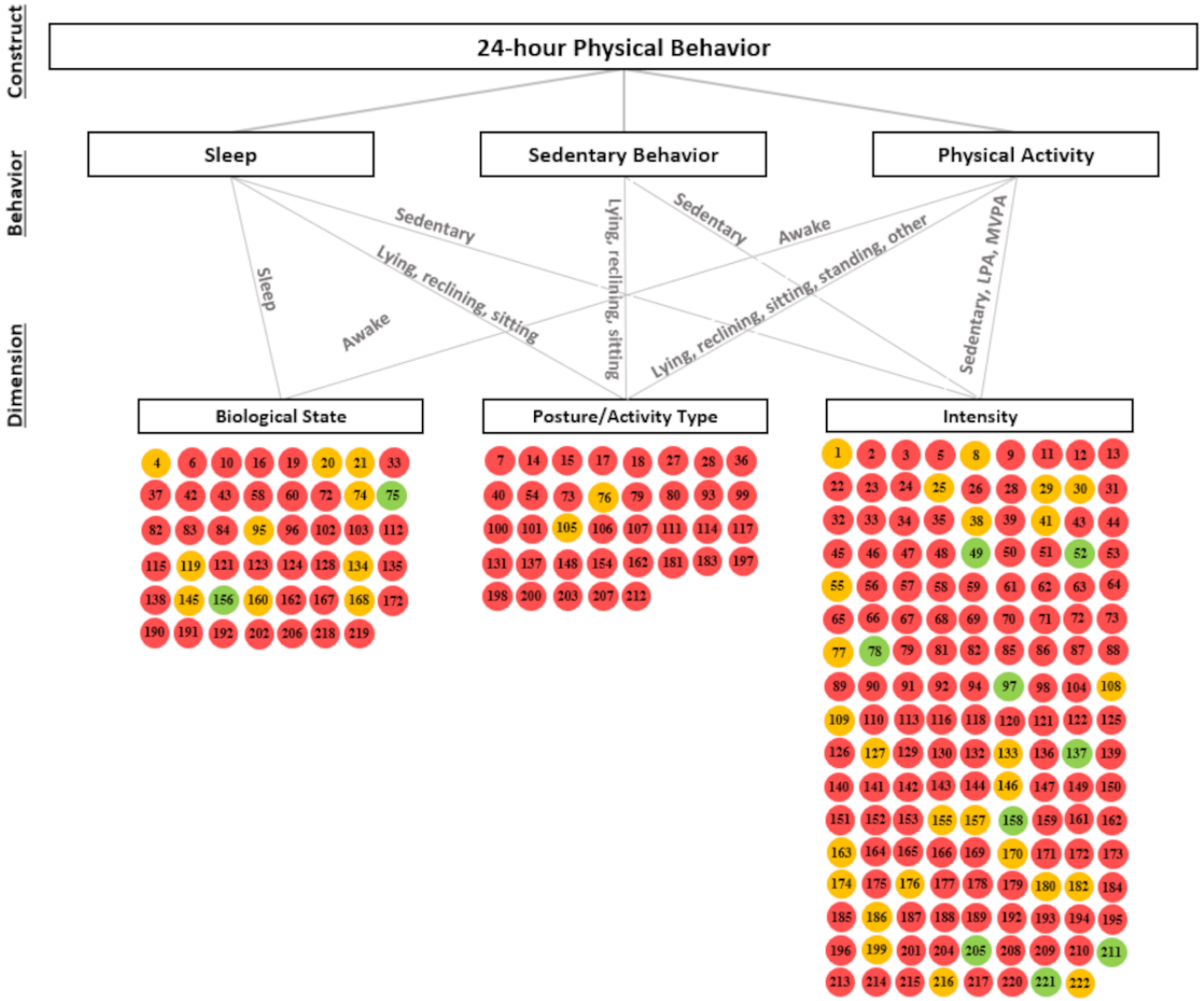
[a]Only relevant for 190 studies.

[b]N/A: not applicable.

[c]Only relevant for 181 studies.

[d]Only relevant for 236 studies.

Figure 2. The overall risk of bias classification separated by different dimensions of the 24-hour physical behavior construct. The number within the circles represents the study number, as listed in the full data extraction. Studies with a green circle were evaluated as low risk, the orange circle represents the quality with some concerns, and red circles represent high risk. LPA: light physical activity; MVPA: moderate to vigorous physical activity.



Discussion

Principal Findings

We evaluated the characteristics and quality of free-living validation studies in which at least one dimension of the 24-hour physical behavior construct (ie, biological state, posture or activity type, and intensity [1,4]) was assessed using wearables and validated against a criterion measure. In summary, the validation of biological state and posture or activity-type outcomes was rare, and almost all of the 163 different types of research- and commercial-grade wearables were validated for only one aspect of the 24-hour physical behavior construct (ie, intensity outcomes). Compared with the selected quality criteria for studies under free-living conditions that are in line with published core principles, recommendations, and expert statements [234-236], most of the reviewed protocols failed to meet the criteria; however, only a few of the evaluated studies were overall ranked with *low risk* of bias or with *some concerns*. Therefore, more high-quality validation studies with adults and older adults under real-life conditions are needed. According to the framework of wearable validation studies [234], the aim of phase 3 studies is to validate device outcomes under real-life conditions against appropriate reference measures.

Criterion Measure

Our evaluation of the most central category physical behavior *criterion measure* followed Keadle et al [234]; for example, physiological outcomes (eg, activity energy expenditure) are recommended to be validated against indirect calorimetry or doubly labeled water, behavioral criterion measures (eg, step count and postures) are recommended to be validated against video-recorded direct observation [234], and the recommended criterion measure for differentiation between sleep and wake patterns is polysomnography [243,244]. Notably, only 40.9% (91/222) of the reviewed studies used the recommended gold standard. Primarily, research-grade devices served as criterion measures, which is highly critical as there is no evidence that wearables can serve as a basis for validating other wearables and offer a high risk of bias regarding criterion validity [233,245,246].

Study Duration

Optimally, study protocols take place over a 24-hour period over multiple days, thus covering a wide range of representative habitual activities [233,234]. This recommended criterion was covered by 2 signaling questions. First, we evaluated whether data collection was not restricted to one particular setting (eg, at work or at home), which was met by nearly all the reviewed studies. Second, as it is almost not feasible to collect data over several days for criterion measures such as video recording [233,234], we specified at least 2 days for a *low-risk* classification. Two-thirds of the reviewed studies met this criterion. However, we identified a considerable number of studies (69/222, 31.1%) that collected data over a short period (≤1 day). The risk of bias might have been present as the setting was restricted to a specific environment (eg, at work), thus limiting the ability to capture a wide range of habitual behaviors. Moreover, reactivity is a serious issue that reveals a potential error source when collecting data from wearables. Researchers expected that reactivity would be a time issue, implying that participants may change their behavior at the beginning of the monitoring period but return to a more stable pattern later [247,248]. Similar effects have been observed in sleep laboratories using polysomnographic monitoring [249].

Study Population

Ideally, the validity of wearables can be generalized to a wide range of diverse samples (eg, age, sex, ethnicity, and health condition) [234,250]. While focusing on adults and older adults (aged ≥18 years), this review revealed that 78.4% (174/222) of the studies included samples between 18 and 64 years of age, whereas there was a lack of studies that included older adults. Critically, most devices (99/163, 60.7%) were validated only once. According to the recommended principle, validation study protocols should include either a variety of cohorts within a single study or a series of studies with different participant characteristics [233,234,251]. For example, we could only identify 20.3% (45/222) of studies that included samples with restricted health conditions. As a practical implication, a given wearable device might be valid for healthy adults and older adults but not for those with health restrictions [232]. A solution might be to recruit a larger sample size, which would enable a higher intersubject variability, or to conduct a series of validation studies with varying participant characteristics. Optimally, sample size calculations ensure adequate power for validation purposes [233,252]. Finally, although challenging because of data protection guidelines, we recommend reporting information about ethnicity (reported in 32/222, 14.4% of studies) whenever possible and providing detailed information about inclusion and exclusion criteria regarding the recruiting process and for statistical analyses.

Wearing Position and Types of Wearables

To enable a comparison between different wearables or wearing positions, researchers may simultaneously collect data from multiple sensors or different wearing positions [234,251]. Most of the reviewed studies did not include multiple wearables (eg, research and commercial grade) and did not capture data from the validated devices at different wearing positions (eg, hip or waist, wrist, and thigh). Depending on the primary outcome of interest, the recommendations of where to place the wearable device may vary. For example, to assess sleep-wake patterns, wrist-worn devices may optimize the recording of small movements that occur at the distal extremities when an individual is supine [245,253]. For example, Fairclough et al [254] reported that wrist placement promotes superior compliance than that of the hip position. In contrast, if researchers are interested in differentiating between body postures (eg, sitting vs standing), the thigh might be the position of choice because of the option of wearing the device under clothing to accurately assess intensity and posture or activity types [4]. However, only 1.8% (4/222) of studies validated posture or activity-type outcomes using thigh-worn devices. Future validation studies are needed with multiple devices and different wearing positions to increase comparability and to inform end users of which device to use and where to place it [250]. In addition, future signal analytical research purposes might be valuable in terms of extracting different outcomes from a single wearing position. Moreover, different types of wearables (ie, pedometers and uniaxial, biaxial, and triaxial accelerometers) have been validated. Researchers should be aware that the different types of devices have different technical requirements. For example, uniaxial accelerometers measure acceleration in 1 direction, whereas triaxial accelerometers measure acceleration in 3 directions. Thus, triaxial accelerometers provide more information, which might be helpful in developing further algorithms.

Synchronization, Transparency, and Statistical Analyses

We evaluated whether the studies reported data synchronization, wear time, the algorithm of the validated outcome, and data analyses. As less than half (75/222, 33.8%) of all included studies reported information about the synchronization process between index and criterion measures, potentially introducing errors and biasing results, we suggest future research endeavors to apply time-stamped solutions such as asking participants to perform 3 vertical jumps at the beginning and the end of the measurement [233]. Following practical consideration when applying wearables [255,256], a large number of studies defined a valid day if ≥10 hours of wear time during waking hours were captured. We set the quality criterion to ≥8 hours per day, revealing that 56% (107/190) of studies considered the wear time criteria for a valid day. Capturing shorter periods may increase the risk of bias as less time is available to assess the data in different settings (eg, at home or at work).

A critical aspect from the perspective of transparency is the presentation of algorithms. Only 43.7% (97/222) of studies reported the algorithm or at least cited further information on the validated outcomes. In particular, no information about the used algorithms was provided in studies in which a commercial-grade device was validated. At this point, researchers often do not have access to the raw data of commercial-grade wearables or the *black box* algorithms. Moreover, companies can update wearables' firmware or algorithms at any time, which hinders comparability [257]. In addition, the pace at which technology is evolving in optimizing algorithms far exceeds the pace of published validation research

[238]. Open-source methods that are more flexible in using algorithms for different devices are needed [233,234].

A quality criterion for the used statistical analyses was not set because of the lack of consistent statistical guidelines for reporting the validity of an activity monitor. Most of the reviewed studies used traditional statistical tests of differences such as *t* tests or ANOVAs. Optimally and in line with recently published suggestions, researchers should integrate different analytical approaches such as combining traditional analyses with equivalence testing and including epoch-by-epoch comparisons whenever possible [234,258].

Limitations

Some points merit further discussion. First, the evaluation of study quality was based on self-selected criteria. In particular, we selected the QUADAS-2 [240] tool and added further signaling questions in line with core principles, recommendations, and expert statements [232-234]. However, as we are not aware of any further quality tools and signaling questions for wearable validation purposes, our selected criteria can serve as a starting point for future systematic reviews that focus on the study quality of wearable technology under free-living conditions. Second, our included validation studies were published between 1987 and 2021. Given the rapid development of wearable technologies and the increasing availability of different research and commercial-grade devices, quality standards have been developed. Thus, when interpreting the study protocols, the time during which the study was conducted should be considered. Third, our review focused on the quality of study protocols. However, we did not take into account further important considerations when using wearables, such as wear or nonwear time algorithms, costs of the monitor, or time of data processing [35,250]. Fourth, our findings were limited to our search strategy; thus, we may have missed further validation studies. However, we applied backward and forward citation searches through the reference lists of the included studies to identify articles that may not have appeared in our search. Finally, this systematic review was limited to articles published in the English language.

Conclusions

Currently, there is a wealth of research on commercial-grade wearables; however, the quality of published validation protocols in adults had not been assessed thus far. However, this is a critical step to enable both researchers and consumers to make guided decisions on which study to rely on and which device to use. To this end, our review unraveled that most validation studies did not meet the recommended quality principles [233,250]. Primarily, there is a lack of validation studies with gold standard reference measures such as video recording, polysomnography, or the doubly labeled water method. Moreover, most devices were validated only once and focused predominantly on intensity measure outcomes. Given the rising interest in the 24-hour physical behavior construct in health research, the next generation of validation studies should consider the validity of >1 aspect of the 24-hour physical behavior construct during a study protocol or to conduct a series of studies. Thus, we conclude that standardized protocols for free-living validation embedded in a framework [234] are urgently needed to inform and guide stakeholders (eg, manufacturers, scientists, and consumers) in (1) selecting wearables for self-tracking purposes, (2) applying wearables in health studies, and (3) fostering innovation to achieve improved validity.

Authors' Contributions

MG, UWEP, MR, and BvHM contributed to the conception and design of the study. MG, RN, DD, and MB contributed to the development of the search strategy. MG, SS, KW, RN, MB, DD, IT, and BvHM conducted the systematic review. MG, KW, SS, IT, DD, MB, and RN performed the data extraction. All authors assisted with the interpretation. MG, UEWP, HB, CRN, BvHM, MR, and KW were the principal authors of the manuscript. All authors contributed to the drafting and revision of the final manuscript. All authors approved the final version of the manuscript.

Conflicts of Interest

UWEP works as a consult for Boehringer-Ingelheim. The authors have no other conflicts to declare.

Multimedia Appendix 1
Characteristics of 24-hour physical behavior, PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist, search terms used in databases, classification tree for the judgment of overall study quality, data extraction, the validity of wearables, an overview of wearables used in validation studies, and risk of bias for the included studies.
[PDF File (Adobe PDF File), 1635 KB - mhealth_v10i6e36377_app1.pdf]

References

1. Rosenberger ME, Fulton JE, Buman MP, Troiano RP, Grandner MA, Buchner DM, et al. The 24-hour activity cycle: a new paradigm for physical activity. Med Sci Sports Exerc 2019 Mar;51(3):454-464 [FREE Full text] [doi: 10.1249/MSS.0000000000001811] [Medline: 30339658]
2. Tremblay MS, Aubert S, Barnes JD, Saunders TJ, Carson V, Latimer-Cheung AE, SBRN Terminology Consensus Project Participants. Sedentary Behavior Research Network (SBRN) - Terminology Consensus Project process and outcome. Int J Behav Nutr Phys Act 2017 Jun 10;14(1):75 [FREE Full text] [doi: 10.1186/s12966-017-0525-8] [Medline: 28599680]

3. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Buman MP, Cardon G, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med 2020 Dec;54(24):1451-1462 [FREE Full text] [doi: 10.1136/bjsports-2020-102955] [Medline: 33239350]
4. Stevens ML, Gupta N, Inan Eroglu E, Crowley PJ, Eroglu B, Bauman A, et al. Thigh-worn accelerometry for measuring movement and posture across the 24-hour cycle: a scoping review and expert statement. BMJ Open Sport Exerc Med 2020 Dec 24;6(1):e000874 [FREE Full text] [doi: 10.1136/bmjsem-2020-000874] [Medline: 33408875]
5. Agogo GO, van der Voet H, Hulshof PJ, Trijsburg L, van Eeuwijk FA, Boshuizen HC. Validation of accelerometer for measuring physical activity in free-living individuals. Balt J Health Phys Act 2018 Mar 31;10(1):7-21. [doi: 10.29359/bjhpa.10.1.01]
6. Albright C, Jerome GJ. The accuracy of talking pedometers when used during free-living: a comparison of four devices. J Vis Impair Blind 2011 May;105(5):299-304. [doi: 10.1177/0145482x1110500507]
7. Alharbi M, Bauman A, Neubeck L, Gallagher R. Validation of Fitbit-Flex as a measure of free-living physical activity in a community-based phase III cardiac rehabilitation population. Eur J Prev Cardiol 2016 Sep;23(14):1476-1485. [doi: 10.1177/2047487316634883] [Medline: 26907794]
8. Ameen MS, Cheung LM, Hauser T, Hahn MA, Schabus M. About the accuracy and problems of consumer devices in the assessment of sleep. Sensors (Basel) 2019 Sep 25;19(19):4160 [FREE Full text] [doi: 10.3390/s19194160] [Medline: 31557952]
9. An HS, Jones GC, Kang SK, Welk GJ, Lee JM. How valid are wearable physical activity trackers for measuring steps? Eur J Sport Sci 2017 Apr;17(3):360-368. [doi: 10.1080/17461391.2016.1255261] [Medline: 27912681]
10. Ancoli-Israel S, Clopton P, Klauber MR, Fell R, Mason W. Use of wrist activity for monitoring sleep/wake in demented nursing-home patients. Sleep 1997 Jan;20(1):24-27 [FREE Full text] [doi: 10.1093/sleep/20.1.24] [Medline: 9130330]
11. Annegarn J, Spruit MA, Uszko-Lencer NH, Vanbelle S, Savelberg HH, Schols AM, et al. Objective physical activity assessment in patients with chronic organ failure: a validation study of a new single-unit activity monitor. Arch Phys Med Rehabil 2011 Nov;92(11):1852-7.e1. [doi: 10.1016/j.apmr.2011.06.021] [Medline: 22032218]
12. Assah FK, Ekelund U, Brage S, Wright A, Mbanya JC, Wareham NJ. Accuracy and validity of a combined heart rate and motion sensor for the measurement of free-living physical activity energy expenditure in adults in Cameroon. Int J Epidemiol 2011 Feb;40(1):112-120. [doi: 10.1093/ije/dyq098] [Medline: 20529884]
13. Au-Yeung WT, Kaye JA, Beattie Z. Step count standardization: validation of step counts from the Withings Activite using PiezoRxD and wGT3X-BT. Annu Int Conf IEEE Eng Med Biol Soc 2020 Jul;2020:4608-4611 [FREE Full text] [doi: 10.1109/EMBC44109.2020.9176511] [Medline: 33019020]
14. Baandrup L, Jennum PJ. A validation of wrist actigraphy against polysomnography in patients with schizophrenia or bipolar disorder. Neuropsychiatr Dis Treat 2015 Aug 28;11:2271-2277 [FREE Full text] [doi: 10.2147/NDT.S88236] [Medline: 26357475]
15. Bai Y, Tompkins C, Gell N, Dione D, Zhang T, Byun W. Comprehensive comparison of Apple Watch and Fitbit monitors in a free-living setting. PLoS One 2021 May 26;16(5):e0251975 [FREE Full text] [doi: 10.1371/journal.pone.0251975] [Medline: 34038458]
16. Barkley JE, Glickman E, Fennell C, Kobak M, Frank M, Farnell G. The validity of the commercially-available, low-cost, wrist-worn Movband accelerometer during treadmill exercise and free-living physical activity. J Sports Sci 2019 Apr;37(7):735-740. [doi: 10.1080/02640414.2018.1523039] [Medline: 30238836]
17. Barone Gibbs B, Paley JL, Jones MA, Whitaker KM, Connolly CP, Catov JM. Validity of self-reported and objectively measured sedentary behavior in pregnancy. BMC Pregnancy Childbirth 2020 Feb 11;20(1):99 [FREE Full text] [doi: 10.1186/s12884-020-2771-z] [Medline: 32046663]
18. Barreira TV, Zderic TW, Schuna Jr JM, Hamilton MT, Tudor-Locke C. Free-living activity counts-derived breaks in sedentary time: are they real transitions from sitting to standing? Gait Posture 2015 Jun;42(1):70-72. [doi: 10.1016/j.gaitpost.2015.04.008] [Medline: 25953504]
19. Bartholdy C, Gudbergsen H, Bliddal H, Kjærgaard M, Lykkegaard KL, Henriksen M. Reliability and construct validity of the SENS Motion® activity measurement system as a tool to detect sedentary behaviour in patients with knee osteoarthritis. Arthritis 2018 Mar 1;2018:6596278 [FREE Full text] [doi: 10.1155/2018/6596278] [Medline: 29686901]
20. Beattie Z, Oyang Y, Statan A, Ghoreyshi A, Pantelopoulos A, Russell A, et al. Estimation of sleep stages in a healthy adult population from optical plethysmography and accelerometer signals. Physiol Meas 2017 Oct 31;38(11):1968-1979. [doi: 10.1088/1361-6579/aa9047] [Medline: 29087960]
21. Berendsen BA, Hendriks MR, Meijer K, Plasqui G, Schaper NC, Savelberg HH. Which activity monitor to use? Validity, reproducibility and user friendliness of three activity monitors. BMC Public Health 2014 Jul 24;14:749 [FREE Full text] [doi: 10.1186/1471-2458-14-749] [Medline: 25059233]
22. Berninger NM, Ten Hoor GA, Plasqui G. Validation of the VitaBit sit-stand tracker: detecting sitting, standing, and activity patterns. Sensors (Basel) 2018 Mar 15;18(3):877 [FREE Full text] [doi: 10.3390/s18030877] [Medline: 29543766]
23. Berryhill S, Morton CJ, Dean A, Berryhill A, Provencio-Dean N, Patel SI, et al. Effect of wearables on sleep in healthy individuals: a randomized crossover trial and validation study. J Clin Sleep Med 2020 May 15;16(5):775-783 [FREE Full text] [doi: 10.5664/jcsm.8356] [Medline: 32043961]

24. Blackwell T, Ancoli-Israel S, Redline S, Stone KL, Osteoporotic Fractures in Men (MrOS) Study Group. Factors that may influence the classification of sleep-wake by wrist actigraphy: the MrOS Sleep Study. J Clin Sleep Med 2011 Aug 15;7(4):357-367 [FREE Full text] [doi: 10.5664/JCSM.1190] [Medline: 21897772]
25. Blackwell T, Redline S, Ancoli-Israel S, Schneider JL, Surovec S, Johnson NL, Study of Osteoporotic Fractures Research Group. Comparison of sleep parameters from actigraphy and polysomnography in older women: the SOF study. Sleep 2008 Feb;31(2):283-291 [FREE Full text] [doi: 10.1093/sleep/31.2.283] [Medline: 18274276]
26. Block VJ, Zhao C, Hollenbach JA, Olgin JE, Marcus GM, Pletcher MJ, et al. Validation of a consumer-grade activity monitor for continuous daily activity monitoring in individuals with multiple sclerosis. Mult Scler J Exp Transl Clin 2019 Nov 21;5(4):2055217319888660 [FREE Full text] [doi: 10.1177/2055217319888660] [Medline: 31803492]
27. Blondeel A, Demeyer H, Janssens W, Troosters T. Accuracy of consumer-based activity trackers as measuring tool and coaching device in patients with COPD and healthy controls. PLoS One 2020 Aug 4;15(8):e0236676 [FREE Full text] [doi: 10.1371/journal.pone.0236676] [Medline: 32750073]
28. Boeselt T, Spielmanns M, Nell C, Storre JH, Windisch W, Magerhans L, et al. Validity and usability of physical activity monitoring in patients with Chronic Obstructive Pulmonary Disease (COPD). PLoS One 2016 Jun 15;11(6):e0157229 [FREE Full text] [doi: 10.1371/journal.pone.0157229] [Medline: 27305105]
29. Bonomi AG, Plasqui G, Goris AH, Westerterp KR. Estimation of free-living energy expenditure using a novel activity monitor designed to minimize obtrusiveness. Obesity (Silver Spring) 2010 Sep;18(9):1845-1851 [FREE Full text] [doi: 10.1038/oby.2010.34] [Medline: 20186133]
30. Bonomi AG, Plasqui G, Goris AH, Westerterp KR. Improving assessment of daily energy expenditure by identifying types of physical activity with a single accelerometer. J Appl Physiol (1985) 2009 Sep;107(3):655-661 [FREE Full text] [doi: 10.1152/japplphysiol.00150.2009] [Medline: 19556460]
31. Bourke AK, Ihlen EA, Helbostad JL. Validation of the activPAL3 in free-living and laboratory scenarios for the measurement of physical activity, stepping, and transitions in older adults. J Meas Phys Behav 2019 Jun;2(2):58-65. [doi: 10.1123/jmpb.2018-0056]
32. Bourke AK, Ihlen EA, Van de Ven P, Nelson J, Helbostad JL. Video analysis validation of a real-time physical activity detection algorithm based on a single waist mounted tri-axial accelerometer sensor. Annu Int Conf IEEE Eng Med Biol Soc 2016 Aug;2016:4881-4884. [doi: 10.1109/EMBC.2016.7591821] [Medline: 28269364]
33. Brage S, Westgate K, Franks PW, Stegle O, Wright A, Ekelund U, et al. Estimation of free-living energy expenditure by heart rate and movement sensing: a doubly-labelled water study. PLoS One 2015 Sep 8;10(9):e0137206 [FREE Full text] [doi: 10.1371/journal.pone.0137206] [Medline: 26349056]
34. Brazeau AS, Beaudoin N, Bélisle V, Messier V, Karelis AD, Rabasa-Lhoret R. Validation and reliability of two activity monitors for energy expenditure assessment. J Sci Med Sport 2016 Jan;19(1):46-50. [doi: 10.1016/j.jsams.2014.11.001] [Medline: 25466490]
35. Brewer W, Swanson BT, Ortiz A. Validity of Fitbit's active minutes as compared with a research-grade accelerometer and self-reported measures. BMJ Open Sport Exerc Med 2017;3(1):e000254 [FREE Full text] [doi: 10.1136/bmjsem-2017-000254] [Medline: 29018543]
36. Breteler MJ, Janssen JH, Spiering W, Kalkman CJ, van Solinge WW, Dohmen DA. Measuring free-living physical activity with three commercially available activity monitors for telemonitoring purposes: Validation study. JMIR Form Res 2019 Apr 24;3(2):e11489 [FREE Full text] [doi: 10.2196/11489] [Medline: 31017587]
37. Brooke SM, An HS, Kang SK, Noble JM, Berg KE, Lee JM. Concurrent validity of wearable activity trackers under free-living conditions. J Strength Cond Res 2017 Apr;31(4):1097-1106. [doi: 10.1519/JSC.0000000000001571] [Medline: 27465631]
38. Brugniaux JV, Niva A, Pulkkinen I, Laukkanen RM, Richalet JP, Pichon AP. Polar Activity Watch 200: a new device to accurately assess energy expenditure. Br J Sports Med 2010 Mar;44(4):245-249. [doi: 10.1136/bjsm.2007.045575] [Medline: 18413338]
39. Busse ME, van Deursen RW, Wiles CM. Real-life step and activity measurement: reliability and validity. J Med Eng Technol 2009;33(1):33-41. [doi: 10.1080/03091900701682606] [Medline: 19116852]
40. Bussmann JB, van de Laar YM, Neeleman MP, Stam HJ. Ambulatory accelerometry to quantify motor behaviour in patients after failed back surgery: a validation study. Pain 1998 Feb;74(2-3):153-161. [doi: 10.1016/s0304-3959(97)00161-9] [Medline: 9520229]
41. Cabanas-Sánchez V, Higueras-Fresnillo S, De la Cámara M, Veiga OL, Martinez-Gomez D. Automated algorithms for detecting sleep period time using a multi-sensor pattern-recognition activity monitor from 24 h free-living data in older adults. Physiol Meas 2018 May 16;39(5):055002. [doi: 10.1088/1361-6579/aabf26] [Medline: 29667936]
42. Calabro MA, Kim Y, Franke WD, Stewart JM, Welk GJ. Objective and subjective measurement of energy expenditure in older adults: a doubly labeled water study. Eur J Clin Nutr 2015 Jul;69(7):850-855 [FREE Full text] [doi: 10.1038/ejcn.2014.241] [Medline: 25351651]
43. Carpenter C, Yang CH, West D. A comparison of sedentary behavior as measured by the Fitbit and ActivPAL in college students. Int J Environ Res Public Health 2021 Apr 08;18(8):3914 [FREE Full text] [doi: 10.3390/ijerph18083914] [Medline: 33917879]

44. Campos C, DePaul VG, Knorr S, Wong JS, Mansfield A, Patterson KK. Validity of the ActiGraph activity monitor for individuals who walk slowly post-stroke. Top Stroke Rehabil 2018 May;25(4):295-304. [doi: 10.1080/10749357.2018.1446487] [Medline: 29557275]
45. Carter J, Wilkinson D, Blacker S, Rayson M, Bilzon J, Izard R, et al. An investigation of a novel three-dimensional activity monitor to predict free-living energy expenditure. J Sports Sci 2008 Apr;26(6):553-561. [doi: 10.1080/02640410701708979] [Medline: 18344125]
46. Castner J, Mammen MJ, Jungquist CR, Licata O, Pender JJ, Wilding GE, et al. Validation of fitness tracker for sleep measures in women with asthma. J Asthma 2019 Jul;56(7):719-730 [FREE Full text] [doi: 10.1080/02770903.2018.1490753] [Medline: 29972657]
47. Cellini N, McDevitt EA, Mednick SC, Buman MP. Free-living cross-comparison of two wearable monitors for sleep and physical activity in healthy young adults. Physiol Behav 2016 Apr 01;157:79-86 [FREE Full text] [doi: 10.1016/j.physbeh.2016.01.034] [Medline: 26821185]
48. Chakravarthy A, Resnick B. Reliability and validity testing of the MotionWatch 8 in older adults. J Nurs Meas 2017 Dec 01;25(3):549-558. [doi: 10.1891/1061-3749.25.3.549] [Medline: 29268834]
49. Choquette S, Chuin A, Lalancette DA, Brochu M, Dionne IJ. Predicting energy expenditure in elders with the metabolic cost of activities. Med Sci Sports Exerc 2009 Oct;41(10):1915-1920. [doi: 10.1249/MSS.0b013e3181a6164a] [Medline: 19727021]
50. Chowdhury EA, Western MJ, Nightingale TE, Peacock OJ, Thompson D. Assessment of laboratory and daily energy expenditure estimates from consumer multi-sensor physical activity monitors. PLoS One 2017 Feb 24;12(2):e0171720 [FREE Full text] [doi: 10.1371/journal.pone.0171720] [Medline: 28234979]
51. Chu AH, Ng SH, Paknezhad M, Gauterin A, Koh D, Brown MS, et al. Comparison of wrist-worn Fitbit Flex and waist-worn ActiGraph for measuring steps in free-living adults. PLoS One 2017 Feb 24;12(2):e0172535 [FREE Full text] [doi: 10.1371/journal.pone.0172535] [Medline: 28234953]
52. Clemes SA, O'Connell S, Rogan LM, Griffiths PL. Evaluation of a commercially available pedometer used to promote physical activity as part of a national programme. Br J Sports Med 2010 Dec;44(16):1178-1183. [doi: 10.1136/bjsm.2009.061085] [Medline: 19474004]
53. Colbert LH, Matthews CE, Havighurst TC, Kim K, Schoeller DA. Comparative validity of physical activity measures in older adults. Med Sci Sports Exerc 2011 May;43(5):867-876 [FREE Full text] [doi: 10.1249/MSS.0b013e3181fc7162] [Medline: 20881882]
54. De Cocker KA, De Meyer J, De Bourdeaudhuij IM, Cardon GM. Non-traditional wearing positions of pedometers: validity and reliability of the Omron HJ-203-ED pedometer under controlled and free-living conditions. J Sci Med Sport 2012 Sep;15(5):418-424. [doi: 10.1016/j.jsams.2012.02.002] [Medline: 22483530]
55. Collins JE, Yang HY, Trentadue TP, Gong Y, Losina E. Validation of the Fitbit Charge 2 compared to the ActiGraph GT3X+ in older adults with knee osteoarthritis in free-living conditions. PLoS One 2019 Jan 30;14(1):e0211231 [FREE Full text] [doi: 10.1371/journal.pone.0211231] [Medline: 30699159]
56. Connolly CP, Dahmen J, Catena RD, Campbell N, Montoye AH. Physical activity monitor accuracy for overground walking and free-living conditions among pregnant women. J Meas Phys Behav 2020 Feb 2;3(2):100-109. [doi: 10.1123/jmpb.2019-0059]
57. Correa JB, Apolzan JW, Shepard DN, Heil DP, Rood JC, Martin CK. Evaluation of the ability of three physical activity monitors to predict weight change and estimate energy expenditure. Appl Physiol Nutr Metab 2016 Jul;41(7):758-766 [FREE Full text] [doi: 10.1139/apnm-2015-0461] [Medline: 27270210]
58. Cuberek R, El Ansari W, Frömel K, Skalik K, Sigmund E. A comparison of two motion sensors for the assessment of free-living physical activity of adolescents. Int J Environ Res Public Health 2010 Apr;7(4):1558-1576 [FREE Full text] [doi: 10.3390/ijerph7041558] [Medline: 20617046]
59. Culhane KM, Lyons GM, Hilton D, Grace PA, Lyons D. Long-term mobility monitoring of older adults using accelerometers in a clinical environment. Clin Rehabil 2004 May;18(3):335-343. [doi: 10.1191/0269215504cr734oa] [Medline: 15137565]
60. Davidson L, McNeill G, Haggarty P, Smith JS, Franklin MF. Free-living energy expenditure of adult men assessed by continuous heart-rate monitoring and doubly-labelled water. Br J Nutr 1997 Nov;78(5):695-708. [doi: 10.1079/bjn19970188] [Medline: 9389894]
61. Degroote L, De Bourdeaudhuij I, Verloigne M, Poppe L, Crombez G. The accuracy of smart devices for measuring physical activity in daily life: validation study. JMIR Mhealth Uhealth 2018 Dec 13;6(12):e10972 [FREE Full text] [doi: 10.2196/10972] [Medline: 30545810]
62. Degroote L, Hamerlinck G, Poels K, Maher C, Crombez G, De Bourdeaudhuij I, et al. Low-Cost consumer-based trackers to measure physical activity and sleep duration among adults in free-living conditions: validation study. JMIR Mhealth Uhealth 2020 May 19;8(5):e16674 [FREE Full text] [doi: 10.2196/16674] [Medline: 32282332]
63. DeShaw KJ, Ellingson L, Bai Y, Lansing J, Perez M, Welk G. Methods for activity monitor validation studies: an example with the Fitbit charge. J Meas Phys Behav 2018;1(3):130-135. [doi: 10.1123/jmpb.2018-0017]

64. Dickinson DL, Cazier J, Cech T. A practical validation study of a commercial accelerometer using good and poor sleepers. Health Psychol Open 2016 Nov 29;3(2):2055102916679012 [FREE Full text] [doi: 10.1177/2055102916679012] [Medline: 28815052]
65. Dominick GM, Winfree KN, Pohlig RT, Papas MA. Physical activity assessment between consumer- and research-grade accelerometers: a comparative study in free-living conditions. JMIR Mhealth Uhealth 2016 Sep 19;4(3):e110 [FREE Full text] [doi: 10.2196/mhealth.6281] [Medline: 27644334]
66. Dieu O, Mikulovic J, Fardy PS, Bui-Xuan G, Béghin L, Vanhelst J. Physical activity using wrist-worn accelerometers: comparison of dominant and non-dominant wrist. Clin Physiol Funct Imaging 2017 Sep;37(5):525-529. [doi: 10.1111/cpf.12337] [Medline: 26749436]
67. Donahoe K, Macdonald DJ, Tremblay MS, Saunders TJ. Validation of PiezoRx pedometer derived sedentary time. Int J Exerc Sci 2018 Jun 1;11(7):552-560 [FREE Full text] [Medline: 29997741]
68. Dondzila CJ, Lewis C, Lopez JR, Parker T. Congruent accuracy of wrist-worn activity trackers during controlled and free-living conditions. Int J Exerc Sci 2018;11(7):575-584.
69. Dondzila CJ, Swartz AM, Miller NE, Lenz EK, Strath SJ. Accuracy of uploadable pedometers in laboratory, overground, and free-living conditions in young and older adults. Int J Behav Nutr Phys Act 2012 Dec 11;9:143 [FREE Full text] [doi: 10.1186/1479-5868-9-143] [Medline: 23232036]
70. Durkalec-Michalski K, Woźniewicz M, Bajerska J, Jeszka J. Comparison of accuracy of various non-calorimetric methods measuring energy expenditure at different intensities. Human Movement 2013;14(2):161-167. [doi: 10.2478/humo-2013-0019]
71. Edwardson CL, Davies M, Khunti K, Yates T, Rowlands AV. Steps per day measured by consumer activity trackers worn at the non-dominant and dominant wrist relative to a waist-worn pedometer. J Meas Phys Behav 2018 Mar;1(1):2-8. [doi: 10.1123/jmpb.2017-0001]
72. Ekelund U, Yngve A, Sjöström M, Westerterp K. Field evaluation of the Computer Science and Application's Inc. Activity monitor during running and skating training in adolescent athletes. Int J Sports Med 2000 Nov;21(8):586-592. [doi: 10.1055/s-2000-8487] [Medline: 11156280]
73. Farina N, Lowry RG. The validity of consumer-level activity monitors in healthy older adults in free-living conditions. J Aging Phys Act 2018 Jan 01;26(1):128-135. [doi: 10.1123/japa.2016-0344] [Medline: 28595019]
74. Feito Y, Bassett DR, Thompson DL. Evaluation of activity monitors in controlled and free-living environments. Med Sci Sports Exerc 2012 Apr;44(4):733-741. [doi: 10.1249/MSS.0b013e3182351913] [Medline: 21904249]
75. Ferguson T, Rowlands AV, Olds T, Maher C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. Int J Behav Nutr Phys Act 2015 Mar 27;12:42 [FREE Full text] [doi: 10.1186/s12966-015-0201-9] [Medline: 25890168]
76. Feito Y, Garner HR, Bassett DR. Evaluation of ActiGraph's low-frequency filter in laboratory and free-living environments. Med Sci Sports Exerc 2015 Jan;47(1):211-217. [doi: 10.1249/MSS.0000000000000395] [Medline: 24870583]
77. Fokkenrood HJ, Verhofstad N, van den Houten MM, Lauret GJ, Wittens C, Scheltinga MR, et al. Physical activity monitoring in patients with peripheral arterial disease: validation of an activity monitor. Eur J Vasc Endovasc Surg 2014 Aug;48(2):194-200 [FREE Full text] [doi: 10.1016/j.ejvs.2014.04.003] [Medline: 24880631]
78. Full KM, Kerr J, Grandner MA, Malhotra A, Moran K, Godoble S, et al. Validation of a physical activity accelerometer device worn on the hip and wrist against polysomnography. Sleep Health 2018 Apr;4(2):209-216 [FREE Full text] [doi: 10.1016/j.sleh.2017.12.007] [Medline: 29555136]
79. Fuller KL, Juliff L, Gore CJ, Peiffer JJ, Halson SL. Software thresholds alter the bias of actigraphy for monitoring sleep in team-sport athletes. J Sci Med Sport 2017 Aug;20(8):756-760. [doi: 10.1016/j.jsams.2016.11.021] [Medline: 28189461]
80. Fullerton E, Heller B, Munoz-Organero M. Recognizing human activity in free-living using multiple body-worn accelerometers. IEEE Sensors J 2017 Aug 15;17(16):5290-5297 [FREE Full text] [doi: 10.1109/jsen.2017.2722105]
81. Gardner AW, Poehlman ET. Assessment of free-living daily physical activity in older claudicants: validation against the doubly labeled water technique. J Gerontol A Biol Sci Med Sci 1998 Jul;53(4):M275-M280. [doi: 10.1093/gerona/53a.4.m275] [Medline: 18314566]
82. Garnotel M, Bastian T, Romero-Ugalde HM, Maire A, Dugas J, Zahariev A, et al. Prior automatic posture and activity identification improves physical activity energy expenditure prediction from hip-worn triaxial accelerometry. J Appl Physiol (1985) 2018 Mar 01;124(3):780-790 [FREE Full text] [doi: 10.1152/japplphysiol.00556.2017] [Medline: 29191980]
83. Gill JM, Hawari NS, Maxwell DJ, Louden D, Mourselas N, Bunn C, et al. Validation of a novel device to measure and provide feedback on sedentary behavior. Med Sci Sports Exerc 2018 Mar;50(3):525-532 [FREE Full text] [doi: 10.1249/MSS.0000000000001458] [Medline: 29040225]
84. Godfrey A, Culhane KM, Lyons GM. Comparison of the performance of the activPAL Professional physical activity logger to a discrete accelerometer-based activity monitor. Med Eng Phys 2007 Oct;29(8):930-934. [doi: 10.1016/j.medengphy.2006.10.001] [Medline: 17134934]
85. Gomersall SR, Ng N, Burton NW, Pavey TG, Gilson ND, Brown WJ. Estimating physical activity and sedentary behavior in a free-living context: a pragmatic comparison of consumer-based activity trackers and actigraph accelerometry. J Med Internet Res 2016 Sep 07;18(9):e239 [FREE Full text] [doi: 10.2196/jmir.5531] [Medline: 27604226]

86. Gruwez A, Libert W, Ameye L, Bruyneel M. Reliability of commercially available sleep and activity trackers with manual switch-to-sleep mode activation in free-living healthy individuals. Int J Med Inform 2017 Jun;102:87-92. [doi: 10.1016/j.ijmedinf.2017.03.008] [Medline: 28495352]
87. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Performance assessment of new-generation Fitbit technology in deriving sleep parameters and stages. Chronobiol Int 2020 Jan;37(1):47-59. [doi: 10.1080/07420528.2019.1682006] [Medline: 31718308]
88. Hargens TA, Deyarmin KN, Snyder KM, Mihalik AG, Sharpe LE. Comparison of wrist-worn and hip-worn activity monitors under free living conditions. J Med Eng Technol 2017 Apr;41(3):200-207. [doi: 10.1080/03091902.2016.1271046] [Medline: 28078908]
89. Hamill K, Jumabhoy R, Kahawage P, de Zambotti M, Walters EM, Drummond SP. Validity, potential clinical utility and comparison of a consumer activity tracker and a research-grade activity tracker in insomnia disorder II: outside the laboratory. J Sleep Res 2020 Feb;29(1):e12944 [FREE Full text] [doi: 10.1111/jsr.12944] [Medline: 31680327]
90. Henriksen A, Grimsgaard S, Horsch A, Hartvigsen G, Hopstock L. Validity of the polar M430 activity monitor in free-living conditions: validation study. JMIR Form Res 2019 Aug 16;3(3):e14438 [FREE Full text] [doi: 10.2196/14438] [Medline: 31420958]
91. Härtel S, Gnam JP, Löffler S, Bös K. Estimation of energy expenditure using accelerometers and activity-based energy models—validation of a new device. Eur Rev Aging Phys Act 2011;8(2):109-114. [doi: 10.1007/s11556-010-0074-5]
92. Hernández-Vicente A, Santos-Lozano A, De Cocker K, Garatachea N. Validation study of Polar V800 accelerometer. Ann Transl Med 2016 Aug;4(15):278 [FREE Full text] [doi: 10.21037/atm.2016.07.16] [Medline: 27570772]
93. Herrmann SD, Hart TL, Lee CD, Ainsworth BE. Evaluation of the MyWellness Key accelerometer. Br J Sports Med 2011 Feb;45(2):109-113. [doi: 10.1136/bjsm.2009.062182] [Medline: 19736173]
94. Hickey A, Del Din S, Rochester L, Godfrey A. Detecting free-living steps and walking bouts: validating an algorithm for macro gait analysis. Physiol Meas 2017 Jan;38(1):N1-15. [doi: 10.1088/1361-6579/38/1/N1] [Medline: 27941238]
95. Hickey A, John D, Sasaki JE, Mavilia M, Freedson P. Validity of activity monitor step detection is related to movement patterns. J Phys Act Health 2016 Feb;13(2):145-153. [doi: 10.1123/jpah.2015-0203] [Medline: 26107045]
96. Höchsmann C, Knaier R, Infanger D, Schmidt-Trucksäss A. Validity of smartphones and activity trackers to measure steps in a free-living setting over three consecutive days. Physiol Meas 2020 Jan 30;41(1):015001 [FREE Full text] [doi: 10.1088/1361-6579/ab635f] [Medline: 31851949]
97. Hui J, Heyden R, Bao T, Accettone N, McBay C, Richardson J, et al. Validity of the Fitbit One for measuring activity in community-dwelling stroke survivors. Physiother Can 2018;70(1):81-89 [FREE Full text] [doi: 10.3138/ptc.2016-40.ep] [Medline: 29434422]
98. Hollewand AM, Spijkerman AG, Bilo HJ, Kleefstra N, Kamsma Y, van Hateren KJ. Validity of an accelerometer-based activity monitor system for measuring physical activity in frail older adults. J Aging Phys Act 2016 Oct;24(4):555-558. [doi: 10.1123/japa.2014-0290] [Medline: 26964560]
99. Jean-Louis G, Kripke DF, Cole RJ, Assmus JD, Langer RD. Sleep detection with an accelerometer actigraph: comparisons with polysomnography. Physiol Behav 2001 Jan;72(1-2):21-28. [doi: 10.1016/s0031-9384(00)00355-3] [Medline: 11239977]
100. Jenkins CA, Tiley LC, Lay I, Hartmann JA, Chan JK, Nicholas CL. Comparing GENEActiv against Actiwatch-2 over seven nights using a common sleep scoring algorithm and device-specific wake thresholds. Behav Sleep Med (forthcoming) 2021 Jun 05:1-11. [doi: 10.1080/15402002.2021.1924175] [Medline: 34096407]
101. Johannsen DL, Calabro MA, Stewart J, Franke W, Rood JC, Welk GJ. Accuracy of armband monitors for measuring daily energy expenditure in healthy adults. Med Sci Sports Exerc 2010 Nov;42(11):2134-2140. [doi: 10.1249/MSS.0b013e3181e0b3ff] [Medline: 20386334]
102. Patrik Johansson H, Rossander-Hulthén L, Slinde F, Ekblom B. Accelerometry combined with heart rate telemetry in the assessment of total energy expenditure. Br J Nutr 2006 Mar;95(3):631-639. [doi: 10.1079/bjn20051527] [Medline: 16512950]
103. Júdice PB, Santos DA, Hamilton MT, Sardinha LB, Silva AM. Validity of GT3X and Actiheart to estimate sedentary time and breaks using ActivPAL as the reference in free-living conditions. Gait Posture 2015 May;41(4):917-922. [doi: 10.1016/j.gaitpost.2015.03.326] [Medline: 25852024]
104. Júdice PB, Teixeira L, Silva AM, Sardinha LB. Accuracy of Actigraph inclinometer to classify free-living postures and motion in adults with overweight and obesity. J Sports Sci 2019 Aug;37(15):1708-1716. [doi: 10.1080/02640414.2019.1586281] [Medline: 30843462]
105. Kanda M, Minakata Y, Matsunaga K, Sugiura H, Hirano T, Koarai A, et al. Validation of the triaxial accelerometer for the evaluation of physical activity in Japanese patients with COPD. Intern Med 2012;51(4):369-375 [FREE Full text] [doi: 10.2169/internalmedicine.51.6441] [Medline: 22333371]
106. Kang SG, Kang JM, Ko KP, Park SC, Mariani S, Weng J. Validity of a commercial wearable sleep tracker in adult insomnia disorder patients and good sleepers. J Psychosom Res 2017 Jun;97:38-44. [doi: 10.1016/j.jpsychores.2017.03.009] [Medline: 28606497]
107. Kawada T. Agreement rates for sleep/wake judgments obtained via accelerometer and sleep diary: a comparison. Behav Res Methods 2008 Nov;40(4):1026-1029. [doi: 10.3758/BRM.40.4.1026] [Medline: 19001393]

108. Keating XD, Liu J, Liu X, Shangguan R, Guan J, Chen L. Validity of Fitbit Charge 2 in controlled college physical education settings. ICHPER-SD J Res 2018;9(2):28-35.
109. Kerr J, Carlson J, Godbole S, Cadmus-Bertram L, Bellettiere J, Hartman S. Improving hip-worn accelerometer estimates of sitting using machine learning methods. Med Sci Sports Exerc 2018 Jul;50(7):1518-1524 [FREE Full text] [doi: 10.1249/MSS.0000000000001578] [Medline: 29443824]
110. Kerr J, Marshall SJ, Godbole S, Chen J, Legge A, Doherty AR, et al. Using the SenseCam to improve classifications of sedentary behavior in free-living settings. Am J Prev Med 2013 Mar;44(3):290-296. [doi: 10.1016/j.amepre.2012.11.004] [Medline: 23415127]
111. Kim Y, Barry VW, Kang M. Validation of the ActiGraph GT3X and activPAL accelerometers for the assessment of sedentary behavior. Meas Phys Educ Exerc Sci 2015 Aug 19;19(3):125-137. [doi: 10.1080/1091367x.2015.1054390]
112. Kinnunen H, Häkkinen K, Schumann M, Karavirta L, Westerterp KR, Kyröläinen H. Training-induced changes in daily energy expenditure: methodological evaluation using wrist-worn accelerometer, heart rate monitor, and doubly labeled water technique. PLoS One 2019 Jul 10;14(7):e0219563 [FREE Full text] [doi: 10.1371/journal.pone.0219563] [Medline: 31291373]
113. Kinnunen H, Tanskanen M, Kyröläinen H, Westerterp KR. Wrist-worn accelerometers in assessment of energy expenditure during intensive training. Physiol Meas 2012 Nov;33(11):1841-1854. [doi: 10.1088/0967-3334/33/11/1841] [Medline: 23110981]
114. Koehler K, Braun H, de Marées M, Fusch G, Fusch C, Schaenzer W. Assessing energy expenditure in male endurance athletes: validity of the SenseWear Armband. Med Sci Sports Exerc 2011 Jul;43(7):1328-1333. [doi: 10.1249/MSS.0b013e31820750f5] [Medline: 21131865]
115. Koenders N, Seeger JP, van der Giessen T, van den Hurk TJ, Smits IG, Tankink AM, et al. Validation of a wireless patch sensor to monitor mobility tested in both an experimental and a hospital setup: a cross-sectional study. PLoS One 2018 Oct 25;13(10):e0206304 [FREE Full text] [doi: 10.1371/journal.pone.0206304] [Medline: 30359448]
116. Kogure T, Shirakawa S, Shimokawa M, Hosokawa Y. Automatic sleep/wake scoring from body motion in bed: validation of a newly developed sensor placed under a mattress. J Physiol Anthropol 2011;30(3):103-109 [FREE Full text] [doi: 10.2114/jpa2.30.103] [Medline: 21636953]
117. Kooiman TJ, Dontje ML, Sprenger SR, Krijnen WP, van der Schans CP, de Groot M. Reliability and validity of ten consumer activity trackers. BMC Sports Sci Med Rehabil 2015 Oct 12;7:24 [FREE Full text] [doi: 10.1186/s13102-015-0018-5] [Medline: 26464801]
118. Kozey-Keadle S, Libertine A, Lyden K, Staudenmayer J, Freedson PS. Validation of wearable monitors for assessing sedentary behavior. Med Sci Sports Exerc 2011 Aug;43(8):1561-1567. [doi: 10.1249/MSS.0b013e31820ce174] [Medline: 21233777]
119. Kubala AG, Barone Gibbs B, Buysse DJ, Patel SR, Hall MH, Kline CE. Field-based measurement of sleep: agreement between six commercial activity monitors and a validated accelerometer. Behav Sleep Med 2020;18(5):637-652 [FREE Full text] [doi: 10.1080/15402002.2019.1651316] [Medline: 31455144]
120. Kumahara H, Ayabe M, Ichibakase M, Tashima A, Chiwata M, Takashi T. Validity of activity monitors worn at multiple nontraditional locations under controlled and free-living conditions in young adult women. Appl Physiol Nutr Metab 2015 May;40(5):448-456. [doi: 10.1139/apnm-2014-0183] [Medline: 25832964]
121. Kurita S, Yano S, Ishii K, Shibata A, Sasai H, Nakata Y, et al. Comparability of activity monitors used in Asian and Western-country studies for assessing free-living sedentary behaviour. PLoS One 2017 Oct 18;12(10):e0186523 [FREE Full text] [doi: 10.1371/journal.pone.0186523] [Medline: 29045441]
122. Kwon S, Wan N, Burns RD, Brusseau TA, Kim Y, Kumar S, et al. The validity of MotionSense HRV in estimating sedentary behavior and physical activity under free-living and simulated activity settings. Sensors (Basel) 2021 Feb 18;21(4):1411 [FREE Full text] [doi: 10.3390/s21041411] [Medline: 33670507]
123. Latshang TD, Mueller DJ, Lo Cascio CM, Stöwhas AC, Stadelmann K, Tesler N, et al. Actigraphy of wrist and ankle for measuring sleep duration in altitude travelers. High Alt Med Biol 2016 Sep;17(3):194-202. [doi: 10.1089/ham.2016.0006] [Medline: 27383065]
124. Le Masurier GC, Lee SM, Tudor-Locke C. Motion sensor accuracy under controlled and free-living conditions. Med Sci Sports Exerc 2004 May;36(5):905-910. [doi: 10.1249/01.mss.0000126777.50188.73] [Medline: 15126728]
125. Lebleu J, Detrembleur C, Guebels C, Hamblenne P, Valet M. Concurrent validity of Nokia Go activity tracker in walking and free-living conditions. J Eval Clin Pract 2020 Feb;26(1):223-228. [doi: 10.1111/jep.13125] [Medline: 30874338]
126. Lee JA, Laurson KR. Validity of the SenseWear armband step count measure during controlled and free-living conditions. J Exerc Sci Fit 2015 Jun;13(1):16-23 [FREE Full text] [doi: 10.1016/j.jesf.2014.11.002] [Medline: 29541094]
127. Lee JA, Williams SM, Brown DD, Laurson KR. Concurrent validation of the Actigraph gt3x+, Polar Active accelerometer, Omron HJ-720 and Yamax Digiwalker SW-701 pedometer step counts in lab-based and free-living settings. J Sports Sci 2015;33(10):991-1000. [doi: 10.1080/02640414.2014.981848] [Medline: 25517396]
128. Lee JM, Byun W, Keill A, Dinkel D, Seo Y. Comparison of wearable trackers' ability to estimate sleep. Int J Environ Res Public Health 2018 Jun 15;15(6):1265 [FREE Full text] [doi: 10.3390/ijerph15061265] [Medline: 29914050]

129. Lee JY, Kwon S, Kim WS, Hahn SJ, Park J, Paik NJ. Feasibility, reliability, and validity of using accelerometers to measure physical activities of patients with stroke during inpatient rehabilitation. PLoS One 2018 Dec 31;13(12):e0209607 [FREE Full text] [doi: 10.1371/journal.pone.0209607] [Medline: 30596694]
130. Lee PH, Suen LK. The convergent validity of Actiwatch 2 and ActiGraph Link accelerometers in measuring total sleeping period, wake after sleep onset, and sleep efficiency in free-living condition. Sleep Breath 2017 Mar;21(1):209-215. [doi: 10.1007/s11325-016-1406-0] [Medline: 27614441]
131. Leenders NY, Sherman WM, Nagaraja HN. Energy expenditure estimated by accelerometry and doubly labeled water: do they agree? Med Sci Sports Exerc 2006 Dec;38(12):2165-2172. [doi: 10.1249/01.mss.0000235883.94357.95] [Medline: 17146325]
132. Li X, Zhang Y, Jiang F, Zhao H. A novel machine learning unsupervised algorithm for sleep/wake identification using actigraphy. Chronobiol Int 2020 Jul;37(7):1002-1015. [doi: 10.1080/07420528.2020.1754848] [Medline: 32342702]
133. Liu S, Brooks D, Thomas S, Eysenbach G, Nolan RP. Lifesource XL-18 pedometer for measuring steps under controlled and free-living conditions. J Sports Sci 2015;33(10):1001-1006. [doi: 10.1080/02640414.2014.981847] [Medline: 25517481]
134. Löf M, Henriksson H, Forsum E. Evaluations of Actiheart, IDEEA® and RT3 monitors for estimating activity energy expenditure in free-living women. J Nutr Sci 2013 Sep 6;2:e31 [FREE Full text] [doi: 10.1017/jns.2013.18] [Medline: 25191581]
135. Lyden K, Keadle SK, Staudenmayer J, Freedson PS. The activPALTM accurately classifies activity intensity categories in healthy adults. Med Sci Sports Exerc 2017 May;49(5):1022-1028 [FREE Full text] [doi: 10.1249/MSS.0000000000001177] [Medline: 28410327]
136. Lyden K, Kozey Keadle SL, Staudenmayer JW, Freedson PS. Validity of two wearable monitors to estimate breaks from sedentary time. Med Sci Sports Exerc 2012 Nov;44(11):2243-2252 [FREE Full text] [doi: 10.1249/MSS.0b013e318260c477] [Medline: 22648343]
137. Mackey DC, Manini TM, Schoeller DA, Koster A, Glynn NW, Goodpaster BH, Health, Aging, and Body Composition Study. Validation of an armband to measure daily energy expenditure in older adults. J Gerontol A Biol Sci Med Sci 2011 Oct;66(10):1108-1113 [FREE Full text] [doi: 10.1093/gerona/glr101] [Medline: 21734231]
138. Madrid-Navarro CJ, Puertas Cuesta FJ, Escamilla-Sevilla F, Campos M, Ruiz Abellán F, Rol MA, et al. Validation of a device for the ambulatory monitoring of sleep patterns: a pilot study on Parkinson's disease. Front Neurol 2019 Apr 11;10:356 [FREE Full text] [doi: 10.3389/fneur.2019.00356] [Medline: 31031690]
139. Mantua J, Gravel N, Spencer RM. Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography. Sensors (Basel) 2016 May 05;16(5):646 [FREE Full text] [doi: 10.3390/s16050646] [Medline: 27164110]
140. Marcotte RT, Petrucci Jr GJ, Cox MF, Freedson PS, Staudenmayer JW, Sirard JR. Estimating sedentary time from a hip- and wrist-worn accelerometer. Med Sci Sports Exerc 2020 Jan;52(1):225-232. [doi: 10.1249/MSS.0000000000002099] [Medline: 31343523]
141. Matthews CE, Kozey Keadle S, Moore SC, Schoeller DS, Carroll RJ, Troiano RP, et al. Measurement of active and sedentary behavior in context of large epidemiologic studies. Med Sci Sports Exerc 2018 Feb;50(2):266-276 [FREE Full text] [doi: 10.1249/MSS.0000000000001428] [Medline: 28930863]
142. McDevitt B, Moore L, Akhtar N, Connolly J, Doherty R, Scott W. Validity of a novel research-grade physical activity and sleep monitor for continuous remote patient monitoring. Sensors (Basel) 2021 Mar 13;21(6):2034 [FREE Full text] [doi: 10.3390/s21062034] [Medline: 33805690]
143. McGinley SK, Armstrong MJ, Khandwala F, Zanuso S, Sigal RJ. Assessment of the MyWellness Key accelerometer in people with type 2 diabetes. Appl Physiol Nutr Metab 2015 Nov;40(11):1193-1198. [doi: 10.1139/apnm-2015-0169] [Medline: 26489052]
144. McVeigh JA, Ellis J, Ross C, Tang K, Wan P, Halse RE, et al. Convergent validity of the Fitbit Charge 2 to measure sedentary behavior and physical activity in overweight and obese adults. J Meas Phys Behav 2021;4(1):39-46. [doi: 10.1123/jmpb.2020-0014]
145. Middelweerd A, Van der Ploeg HP, Van Halteren A, Twisk JW, Brug J, Te Velde SJ. A validation study of the Fitbit one in daily life using different time intervals. Med Sci Sports Exerc 2017 Jun;49(6):1270-1279. [doi: 10.1249/MSS.0000000000001225] [Medline: 28511193]
146. Mikkelsen ML, Berg-Beckhoff G, Frederiksen P, Horgan G, O'Driscoll R, Palmeira AL, et al. Estimating physical activity and sedentary behaviour in a free-living environment: a comparative study between Fitbit Charge 2 and Actigraph GT3X. PLoS One 2020 Jun 11;15(6):e0234426 [FREE Full text] [doi: 10.1371/journal.pone.0234426] [Medline: 32525912]
147. Miyamoto S, Minakata Y, Azuma Y, Kawabe K, Ono H, Yanagimoto R, et al. Verification of a motion sensor for evaluating physical activity in COPD patients. Can Respir J 2018 Apr 23;2018:8343705 [FREE Full text] [doi: 10.1155/2018/8343705] [Medline: 29849834]
148. Moore SA, Hallsworth K, Bluck LJ, Ford GA, Rochester L, Trenell MI. Measuring energy expenditure after stroke: validation of a portable device. Stroke 2012 Jun;43(6):1660-1662. [doi: 10.1161/STROKEAHA.111.646257] [Medline: 22499575]
149. Mouritzen NJ, Larsen LH, Lauritzen MH, Kjær TW. Assessing the performance of a commercial multisensory sleep tracker. PLoS One 2020 Dec 11;15(12):e0243214 [FREE Full text] [doi: 10.1371/journal.pone.0243214] [Medline: 33306678]

150. Murakami H, Kawakami R, Nakae S, Yamada Y, Nakata Y, Ohkawara K, et al. Accuracy of 12 wearable devices for estimating physical activity energy expenditure using a metabolic chamber and the doubly labeled water method: validation study. JMIR Mhealth Uhealth 2019 Aug 02;7(8):e13938 [FREE Full text] [doi: 10.2196/13938] [Medline: 31376273]
151. Narayanan A, Stewart T, Mackay L. A dual-accelerometer system for detecting human movement in a free-living environment. Med Sci Sports Exerc 2020 Jan;52(1):252-258. [doi: 10.1249/MSS.0000000000002107] [Medline: 31361712]
152. Myers J, Dupain M, Vu A, Jaffe A, Smith K, Fonda H, et al. Agreement between activity-monitoring devices during home rehabilitation: a substudy of the AAA STOP trial. J Aging Phys Act 2014 Jan;22(1):87-95. [doi: 10.1123/japa.2012-0133] [Medline: 23416349]
153. Nguyen DM, Lecoultre V, Sunami Y, Schutz Y. Assessment of physical activity and energy expenditure by GPS combined with accelerometry in real-life conditions. J Phys Act Health 2013 Aug;10(6):880-888. [doi: 10.1123/jpah.10.6.880] [Medline: 23072762]
154. O'Brien CM, Duda JL, Kitas GD, Veldhuijzen van Zanten JJ, Metsios GS, Fenton SA. Measurement of sedentary time and physical activity in rheumatoid arthritis: an ActiGraph and activPAL™ validation study. Rheumatol Int 2020 Sep;40(9):1509-1518 [FREE Full text] [doi: 10.1007/s00296-020-04608-2] [Medline: 32472303]
155. O'Neill B, McDonough SM, Wilson JJ, Bradbury I, Hayes K, Kirk A, et al. Comparing accelerometer, pedometer and a questionnaire for measuring physical activity in bronchiectasis: a validity and feasibility study? Respir Res 2017 Jan 14;18(1):16 [FREE Full text] [doi: 10.1186/s12931-016-0497-2] [Medline: 28088206]
156. O'Brien MW, Wojcik WR, D'Entremont L, Fowles JR. Validation of the PiezoRx® step count and moderate to vigorous physical activity times in free living conditions in adults: a pilot study. Int J Exerc Sci 2018 Jan 2;11(7):541-551 [FREE Full text] [Medline: 29541342]
157. Paul SS, Tiedemann A, Hassett LM, Ramsay E, Kirkham C, Chagpar S, et al. Validity of the Fitbit activity tracker for measuring steps in community-dwelling older adults. BMJ Open Sport Exerc Med 2015 Jul 8;1(1):e000013 [FREE Full text] [doi: 10.1136/bmjsem-2015-000013] [Medline: 27900119]
158. Pavey TG, Gomersall SR, Clark BK, Brown WJ. The validity of the GENEActiv wrist-worn accelerometer for measuring adult sedentary time in free living. J Sci Med Sport 2016 May;19(5):395-399. [doi: 10.1016/j.jsams.2015.04.007] [Medline: 25956687]
159. Pomeroy J, Brage S, Curtis JM, Swan PD, Knowler WC, Franks PW. Between-monitor differences in step counts are related to body size: implications for objective physical activity measurement. PLoS One 2011 Apr 27;6(4):e18942 [FREE Full text] [doi: 10.1371/journal.pone.0018942] [Medline: 21556140]
160. Quante M, Kaplan ER, Cailler M, Rueschman M, Wang R, Weng J, et al. Actigraphy-based sleep estimation in adolescents and adults: a comparison with polysomnography using two scoring algorithms. Nat Sci Sleep 2018 Jan 18;10:13-20 [FREE Full text] [doi: 10.2147/NSS.S151085] [Medline: 29403321]
161. Rabinovich RA, Louvaris Z, Raste Y, Langer D, Van Remoortel H, Giavedoni S, PROactive Consortium. Validity of physical activity monitors during daily life in patients with COPD. Eur Respir J 2013 Nov;42(5):1205-1215 [FREE Full text] [doi: 10.1183/09031936.00134312] [Medline: 23397303]
162. Rafamantanantsoa HH, Ebine N, Yoshioka M, Higuchi H, Yoshitake Y, Tanaka H, et al. Validation of three alternative methods to measure total energy expenditure against the doubly labeled water method for older Japanese men. J Nutr Sci Vitaminol (Tokyo) 2002 Dec;48(6):517-523. [doi: 10.3177/jnsv.48.517] [Medline: 12778890]
163. Redenius N, Kim Y, Byun W. Concurrent validity of the Fitbit for assessing sedentary behavior and moderate-to-vigorous physical activity. BMC Med Res Methodol 2019 Feb 07;19(1):29 [FREE Full text] [doi: 10.1186/s12874-019-0668-1] [Medline: 30732582]
164. Regalia G, Gerboni G, Migliorini M, Lai M, Pham J, Puri N, et al. Sleep assessment by means of a wrist actigraphy-based algorithm: agreement with polysomnography in an ambulatory study on older adults. Chronobiol Int 2021 Mar;38(3):400-414. [doi: 10.1080/07420528.2020.1835942] [Medline: 33213222]
165. Reid RE, Insogna JA, Carver TE, Comptour AM, Bewski NA, Sciortino C, et al. Validity and reliability of Fitbit activity monitors compared to ActiGraph GT3X+ with female adults in a free-living environment. J Sci Med Sport 2017 Jun;20(6):578-582. [doi: 10.1016/j.jsams.2016.10.015] [Medline: 27887786]
166. Rosenberger ME, Buman MP, Haskell WL, McConnell MV, Carstensen LL. Twenty-four hours of sleep, sedentary behavior, and physical activity with nine wearable devices. Med Sci Sports Exerc 2016 Mar;48(3):457-465 [FREE Full text] [doi: 10.1249/MSS.0000000000000778] [Medline: 26484953]
167. Rothney MP, Brychta RJ, Meade NN, Chen KY, Buchowski MS. Validation of the ActiGraph two-regression model for predicting energy expenditure. Med Sci Sports Exerc 2010 Sep;42(9):1785-1792 [FREE Full text] [doi: 10.1249/MSS.0b013e3181d5a984] [Medline: 20142778]
168. Rousset S, Fardet A, Lacomme P, Normand S, Montaurier C, Boirie Y, et al. Comparison of total energy expenditure assessed by two devices in controlled and free-living conditions. Eur J Sport Sci 2015;15(5):391-399. [doi: 10.1080/17461391.2014.949309] [Medline: 25141769]
169. Rozanski GM, Aqui A, Sivakumaran S, Mansfield A. Consumer wearable devices for activity monitoring among individuals after a stroke: a prospective comparison. JMIR Cardio 2018 Jan 04;2(1):e1 [FREE Full text] [doi: 10.2196/cardio.8199] [Medline: 31758760]

170. Rutgers CJ, Klijn MJ, Deurenberg P. The assessment of 24-hour energy expenditure in elderly women by minute-by-minute heart rate monitoring. Ann Nutr Metab 1997;41(2):83-88. [doi: 10.1159/000177982] [Medline: 9267582]
171. Sánchez-Ortuño MM, Edinger JD, Means MK, Almirall D. Home is where sleep is: an ecological approach to test the validity of actigraphy for the assessment of insomnia. J Clin Sleep Med 2010 Feb 15;6(1):21-29 [FREE Full text] [Medline: 20191934]
172. Sargent C, Lastella M, Halson SL, Roach GD. The validity of activity monitors for measuring sleep in elite athletes. J Sci Med Sport 2016 Oct;19(10):848-853. [doi: 10.1016/j.jsams.2015.12.007] [Medline: 26794719]
173. Sasaki S, Ukawa S, Okada E, Wenjing Z, Kishi T, Sakamoto A, et al. Comparison of a new wrist-worn accelerometer with a commonly used triaxial accelerometer under free-living conditions. BMC Res Notes 2018 Oct 20;11(1):746 [FREE Full text] [doi: 10.1186/s13104-018-3849-9] [Medline: 30342547]
174. Schmal H, Holsgaard-Larsen A, Izadpanah K, Brønd JC, Madsen CF, Lauritsen J. Validation of activity tracking procedures in elderly patients after operative treatment of proximal femur fractures. Rehabil Res Pract 2018 Jun 19;2018:3521271 [FREE Full text] [doi: 10.1155/2018/3521271] [Medline: 30018824]
175. Schneider PL, Crouter SE, Bassett DR. Pedometer measures of free-living physical activity: comparison of 13 models. Med Sci Sports Exerc 2004 Feb;36(2):331-335. [doi: 10.1249/01.MSS.0000113486.60548.E9] [Medline: 14767259]
176. Scott J, Grierson A, Gehue L, Kallestad H, MacMillan I, Hickie I. Can consumer grade activity devices replace research grade actiwatches in youth mental health settings? Sleep Biol Rhythms 2019 Jan 17;17(2):223-232. [doi: 10.1007/s41105-018-00204-x]
177. Semanik P, Lee J, Pellegrini CA, Song J, Dunlop DD, Chang RW. Comparison of physical activity measures derived from the Fitbit Flex and the ActiGraph GT3X+ in an employee population with chronic knee symptoms. ACR Open Rheumatol 2020 Jan;2(1):48-52 [FREE Full text] [doi: 10.1002/acr2.11099] [Medline: 31943970]
178. Siddall AG, Powell SD, Needham-Beck SC, Edwards VC, Thompson JE, Kefyalew SS, et al. Validity of energy expenditure estimation methods during 10 days of military training. Scand J Med Sci Sports 2019 Sep;29(9):1313-1321. [doi: 10.1111/sms.13488] [Medline: 31136027]
179. Silcott NA, Bassett Jr DR, Thompson DL, Fitzhugh EC, Steeves JA. Evaluation of the Omron HJ-720ITC pedometer under free-living conditions. Med Sci Sports Exerc 2011 Sep;43(9):1791-1797. [doi: 10.1249/MSS.0b013e318212888c] [Medline: 21311356]
180. Silva AM, Santos DA, Matias CN, Júdice PB, Magalhães JP, Ekelund U, et al. Accuracy of a combined heart rate and motion sensor for assessing energy expenditure in free-living adults during a double-blind crossover caffeine trial using doubly labeled water as the reference method. Eur J Clin Nutr 2015 Jan;69(1):20-27. [doi: 10.1038/ejcn.2014.51] [Medline: 24690589]
181. Silva GS, Yang H, Collins JE, Losina E. Validating Fitbit for evaluation of physical activity in patients with knee osteoarthritis: do thresholds matter? ACR Open Rheumatol 2019 Nov;1(9):585-592 [FREE Full text] [doi: 10.1002/acr2.11080] [Medline: 31777843]
182. Šimůnek A, Dygrýn J, Gába A, Jakubec L, Stelzer J, Chmelík F. Validity of Garmin Vívofit and polar loop for measuring daily step counts in free-living conditions in adults. Acta Gymnica 2016 Sep 30;46(3):129-135. [doi: 10.5507/ag.2016.014]
183. Sjöberg V, Westergren J, Monnier A, Lo Martire R, Hagströmer M, Äng BO, et al. Wrist-worn activity trackers in laboratory and free-living settings for patients with chronic pain: criterion validity study. JMIR Mhealth Uhealth 2021 Jan 12;9(1):e24806 [FREE Full text] [doi: 10.2196/24806] [Medline: 33433391]
184. Skipworth RJ, Stene GB, Dahele M, Hendry PO, Small AC, Blum D, European Palliative Care Research Collaborative (EPCRC). Patient-focused endpoints in advanced cancer: criterion-based validation of accelerometer-based activity monitoring. Clin Nutr 2011 Dec;30(6):812-821. [doi: 10.1016/j.clnu.2011.05.010] [Medline: 21733604]
185. Skotte J, Korshøj M, Kristiansen J, Hanisch C, Holtermann A. Detection of physical activity types using triaxial accelerometers. J Phys Act Health 2014 Jan;11(1):76-84. [doi: 10.1123/jpah.2011-0347] [Medline: 23249722]
186. Slinde F, Bertz F, Winkvist A, Ellegård L, Olausson H, Brekke HK. Energy expenditure by multisensor armband in overweight and obese lactating women validated by doubly labeled water. Obesity (Silver Spring) 2013 Nov;21(11):2231-2235 [FREE Full text] [doi: 10.1002/oby.20363] [Medline: 23512821]
187. Smits EJ, Winkler EA, Healy GN, Dall PM, Granat MH, Hodges PW. Comparison of single- and dual-monitor approaches to differentiate sitting from lying in free-living conditions. Scand J Med Sci Sports 2018 Aug;28(8):1888-1896. [doi: 10.1111/sms.13203] [Medline: 29701887]
188. Stein AD, Rivera JM, Pivarnik JM. Measuring energy expenditure in habitually active and sedentary pregnant women. Med Sci Sports Exerc 2003 Aug;35(8):1441-1446. [doi: 10.1249/01.MSS.0000079107.04349.9A] [Medline: 12900702]
189. St-Laurent A, Mony MM, Mathieu M, Ruchat SM. Validation of the Fitbit Zip and Fitbit Flex with pregnant women in free-living conditions. J Med Eng Technol 2018 May;42(4):259-264. [doi: 10.1080/03091902.2018.1472822] [Medline: 30198806]
190. St-Onge M, Mignault D, Allison DB, Rabasa-Lhoret R. Evaluation of a portable device to measure daily energy expenditure in free-living adults. Am J Clin Nutr 2007 Mar;85(3):742-749. [doi: 10.1093/ajcn/85.3.742] [Medline: 17344495]
191. Strath SJ, Brage S, Ekelund U. Integration of physiological and accelerometer data to improve physical activity assessment. Med Sci Sports Exerc 2005 Nov;37(11 Suppl):S563-S571. [doi: 10.1249/01.mss.0000185650.68232.3f] [Medline: 16294119]

192. Sugino A, Minakata Y, Kanda M, Akamatsu K, Koarai A, Hirano T, et al. Validation of a compact motion sensor for the measurement of physical activity in patients with chronic obstructive pulmonary disease. Respiration 2012;83(4):300-307 [FREE Full text] [doi: 10.1159/000330046] [Medline: 21912085]
193. Sushames A, Edwards A, Thompson F, McDermott R, Gebel K. Validity and reliability of Fitbit Flex for step count, moderate to vigorous physical activity and activity energy expenditure. PLoS One 2016 Sep 2;11(9):e0161224 [FREE Full text] [doi: 10.1371/journal.pone.0161224] [Medline: 27589592]
194. Svensson T, Chung UI, Tokuno S, Nakamura M, Svensson AK. A validation study of a consumer wearable sleep tracker compared to a portable EEG system in naturalistic conditions. J Psychosom Res 2019 Nov;126:109822. [doi: 10.1016/j.jpsychores.2019.109822] [Medline: 31499232]
195. te Lindert BH, Van Someren EJ. Sleep estimates using microelectromechanical systems (MEMS). Sleep 2013 May 01;36(5):781-789 [FREE Full text] [doi: 10.5665/sleep.2648] [Medline: 23633761]
196. Tedesco S, Sica M, Ancillao A, Timmons S, Barton J, O'Flynn B. Validity evaluation of the Fitbit Charge2 and the Garmin vivosmart HR+ in free-living environments in an older adult cohort. JMIR Mhealth Uhealth 2019 Jun 19;7(6):e13084 [FREE Full text] [doi: 10.2196/13084] [Medline: 31219048]
197. Toth LP, Park S, Springer CM, Feyerabend MD, Steeves JA, Bassett DR. Video-recorded validation of wearable step counters under free-living conditions. Med Sci Sports Exerc 2018 Jun;50(6):1315-1322. [doi: 10.1249/MSS.0000000000001569] [Medline: 29381649]
198. Thorup CB, Andreasen JJ, Sørensen EE, Grønkjær M, Dinesen BI, Hansen J. Accuracy of a step counter during treadmill and daily life walking by healthy adults and patients with cardiac disease. BMJ Open 2017 Mar 31;7(3):e011742 [FREE Full text] [doi: 10.1136/bmjopen-2016-011742] [Medline: 28363918]
199. Tudor-Locke C, Sisson SB, Lee SM, Craig CL, Plotnikoff RC, Bauman A. Evaluation of quality of commercial pedometers. Can J Public Health 2006;97 Suppl 1:S10-S16 [FREE Full text] [doi: 10.1007/BF03405359] [Medline: 16676833]
200. Tully MA, McBride C, Heron L, Hunter RF. The validation of Fibit Zip™ physical activity monitor as a measure of free-living physical activity. BMC Res Notes 2014 Dec 23;7:952 [FREE Full text] [doi: 10.1186/1756-0500-7-952] [Medline: 25539733]
201. Uiterwaal M, Glerum EB, Busser HJ, van Lummel RC. Ambulatory monitoring of physical activity in working situations, a validation study. J Med Eng Technol 1998;22(4):168-172. [doi: 10.3109/03091909809032535] [Medline: 9680600]
202. Vähä-Ypyä H, Husu P, Suni J, Vasankari T, Sievänen H. Reliable recognition of lying, sitting, and standing with a hip-worn accelerometer. Scand J Med Sci Sports 2018 Mar;28(3):1092-1102. [doi: 10.1111/sms.13017] [Medline: 29144567]
203. Valenti G, Camps SG, Verhoef SP, Bonomi AG, Westerterp KR. Validating measures of free-living physical activity in overweight and obese subjects using an accelerometer. Int J Obes (Lond) 2014 Jul;38(7):1011-1014. [doi: 10.1038/ijo.2013.195] [Medline: 24166066]
204. van Alphen HJ, Waninge A, Minnaert AE, Post WJ, van der Putten AA. Construct validity of the Actiwatch-2 for assessing movement in people with profound intellectual and multiple disabilities. J Appl Res Intellect Disabil 2021 Jan;34(1):99-110 [FREE Full text] [doi: 10.1111/jar.12789] [Medline: 32725667]
205. Van Blarigan EL, Kenfield SA, Tantum L, Cadmus-Bertram LA, Carroll PR, Chan JM. The Fitbit One physical activity tracker in men with prostate cancer: validation study. JMIR Cancer 2017 Apr 18;3(1):e5 [FREE Full text] [doi: 10.2196/cancer.6935] [Medline: 28420602]
206. van de Wouw E, Evenhuis HM, Echteld MA. Comparison of two types of Actiwatch with polysomnography in older adults with intellectual disability: a pilot study. J Intellect Dev Disabil 2013 Sep;38(3):265-273. [doi: 10.3109/13668250.2013.816274] [Medline: 23984885]
207. van den Berg-Emons HJ, Bussmann JB, Balk AH, Stam HJ. Validity of ambulatory accelerometry to quantify physical activity in heart failure. Scand J Rehabil Med 2000 Dec;32(4):187-192. [doi: 10.1080/003655000750060940] [Medline: 11201626]
208. van der Weegen S, Essers H, Spreeuwenberg M, Verwey R, Tange H, de Witte L, et al. Concurrent validity of the MOX activity monitor compared to the ActiGraph GT3X. Telemed J E Health 2015 Apr;21(4):259-266. [doi: 10.1089/tmj.2014.0097] [Medline: 25614938]
209. van Hees VT, Renström F, Wright A, Gradmark A, Catt M, Chen KY, et al. Estimation of daily energy expenditure in pregnant and non-pregnant women using a wrist-worn tri-axial accelerometer. PLoS One 2011;6(7):e22922 [FREE Full text] [doi: 10.1371/journal.pone.0022922] [Medline: 21829556]
210. van Hees VT, Sabia S, Jones SE, Wood AR, Anderson KN, Kivimäki M, et al. Estimating sleep parameters using an accelerometer without sleep diary. Sci Rep 2018 Aug 28;8(1):12975 [FREE Full text] [doi: 10.1038/s41598-018-31266-z] [Medline: 30154500]
211. van Nassau F, Chau JY, Lakerveld J, Bauman AE, van der Ploeg HP. Validity and responsiveness of four measures of occupational sitting and standing. Int J Behav Nutr Phys Act 2015 Nov 25;12:144 [FREE Full text] [doi: 10.1186/s12966-015-0306-1] [Medline: 26608219]
212. Vanhelst J, Mikulovic J, Bui-Xuan G, Dieu O, Blondeau T, Fardy P, et al. Comparison of two ActiGraph accelerometer generations in the assessment of physical activity in free living conditions. BMC Res Notes 2012 Apr 25;5:187 [FREE Full text] [doi: 10.1186/1756-0500-5-187] [Medline: 22534207]

213. Varela Mato V, Yates T, Stensel D, Biddle S, Clemes SA. Concurrent validity of Actigraph-determined sedentary time against the Activpal under free-living conditions in a sample of bus drivers. Meas Phys Educ Exerc Sci 2017 Jun 22;21(4):212-222. [doi: 10.1080/1091367x.2017.1335204]
214. Vetrovsky T, Siranec M, Marencakova J, Tufano JJ, Capek V, Bunc V, et al. Validity of six consumer-level activity monitors for measuring steps in patients with chronic heart failure. PLoS One 2019 Sep 13;14(9):e0222569 [FREE Full text] [doi: 10.1371/journal.pone.0222569] [Medline: 31518367]
215. Villars C, Bergouignan A, Dugas J, Antoun E, Schoeller DA, Roth H, et al. Validity of combining heart rate and uniaxial acceleration to measure free-living physical activity energy expenditure in young men. J Appl Physiol (1985) 2012 Dec 01;113(11):1763-1771 [FREE Full text] [doi: 10.1152/japplphysiol.01413.2011] [Medline: 23019315]
216. Washburn RA, Janney CA, Fenster JR. The validity of objective physical activity monitoring in older individuals. Res Q Exerc Sport 1990 Mar;61(1):114-117. [doi: 10.1080/02701367.1990.10607489] [Medline: 2091162]
217. Webber SC, St John PD. Comparison of ActiGraph GT3X+ and StepWatch step count accuracy in geriatric rehabilitation patients. J Aging Phys Act 2016 Jul;24(3):451-458. [doi: 10.1123/japa.2015-0234] [Medline: 26751505]
218. Welk GJ, McClain JJ, Eisenmann JC, Wickel EE. Field validation of the MTI Actigraph and BodyMedia armband monitor using the IDEEA monitor. Obesity (Silver Spring) 2007 Apr;15(4):918-928 [FREE Full text] [doi: 10.1038/oby.2007.624] [Medline: 17426327]
219. Welk G, Kim Y, Shook RP, Ellingson L, Lobelo RL. Validation of a noninvasive, disposable activity monitor for clinical applications. J Phys Act Health 2017 Jul;14(7):546-551. [doi: 10.1123/jpah.2016-0003] [Medline: 28290750]
220. White T, Westgate K, Hollidge S, Venables M, Olivier P, Wareham N, et al. Estimating energy expenditure from wrist and thigh accelerometry in free-living adults: a doubly labelled water study. Int J Obes (Lond) 2019 Nov;43(11):2333-2342 [FREE Full text] [doi: 10.1038/s41366-019-0352-x] [Medline: 30940917]
221. Whybrow S, Ritz P, Horgan GW, Stubbs RJ. An evaluation of the IDEEA™ activity monitor for estimating energy expenditure. Br J Nutr 2013 Jan 14;109(1):173-183. [doi: 10.1017/S0007114512000645] [Medline: 22464547]
222. Williams JM, Taylor DJ, Slavish DC, Gardner CE, Zimmerman MR, Patel K, et al. Validity of actigraphy in young adults with insomnia. Behav Sleep Med 2020;18(1):91-106. [doi: 10.1080/15402002.2018.1545653] [Medline: 30472879]
223. Winkler EA, Bodicoat DH, Healy GN, Bakrania K, Yates T, Owen N, et al. Identifying adults' valid waking wear time by automated estimation in activPAL data collected with a 24 h wear protocol. Physiol Meas 2016 Oct;37(10):1653-1668. [doi: 10.1088/0967-3334/37/10/1653] [Medline: 27652827]
224. Wu Y, Johns JA, Poitras J, Kimmerly DS, O'Brien MW. Improving the criterion validity of the activPAL in determining physical activity intensity during laboratory and free-living conditions. J Sports Sci 2021 Apr;39(7):826-834. [doi: 10.1080/02640414.2020.1847503] [Medline: 33203323]
225. Yoshida A, Ishikawa-Takata K, Tanaka S, Suzuki N, Nakae S, Murata H, et al. Validity of combination use of activity record and accelerometry to measure free-living total energy expenditure in female endurance runners. J Strength Cond Res 2019 Nov;33(11):2962-2970. [doi: 10.1519/JSC.0000000000002205] [Medline: 29389693]
226. Yamada Y, Hashii-Arishima Y, Yokoyama K, Itoi A, Adachi T, Kimura M. Validity of a triaxial accelerometer and simplified physical activity record in older adults aged 64-96 years: a doubly labeled water study. Eur J Appl Physiol 2018 Oct;118(10):2133-2146. [doi: 10.1007/s00421-018-3944-6] [Medline: 30019086]
227. Thompson WR. Worldwide survey of fitness trends for 2022. ACSMs Health Fit J 2022;26(1):11-20. [doi: 10.1249/FIT.0000000000000732]
228. Troiano RP, Stamatakis E, Bull FC. How can global physical activity surveillance adapt to evolving physical activity guidelines? Needs, challenges and future directions. Br J Sports Med 2020 Dec;54(24):1468-1473 [FREE Full text] [doi: 10.1136/bjsports-2020-102621] [Medline: 33239352]
229. Trost SG. Population-level physical activity surveillance in young people: are accelerometer-based measures ready for prime time? Int J Behav Nutr Phys Act 2020 Mar 18;17(1):28 [FREE Full text] [doi: 10.1186/s12966-020-00929-4] [Medline: 32183807]
230. Burchartz A, Anedda B, Auerswald T, Giurgiu M, Hill H, Ketelhut S, et al. Assessing physical behavior through accelerometry – state of the science, best practices and future directions. Psychol Sport Exerc 2020 Jul;49:101703 [FREE Full text] [doi: 10.1016/j.psychsport.2020.101703]
231. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010 May;19(4):539-549 [FREE Full text] [doi: 10.1007/s11136-010-9606-8] [Medline: 20169472]
232. Bassett Jr DR, Rowlands A, Trost SG. Calibration and validation of wearable monitors. Med Sci Sports Exerc 2012 Jan;44(1 Suppl 1):S32-S38 [FREE Full text] [doi: 10.1249/MSS.0b013e3182399cf7] [Medline: 22157772]
233. Johnston W, Judice PB, Molina García P, Mühlen JM, Lykke Skovgaard E, Stang J, et al. Recommendations for determining the validity of consumer wearable and smartphone step count: expert statement and checklist of the INTERLIVE network. Br J Sports Med 2021 Jul;55(14):780-793 [FREE Full text] [doi: 10.1136/bjsports-2020-103147] [Medline: 33361276]
234. Keadle SK, Lyden KA, Strath SJ, Staudenmayer JW, Freedson PS. A framework to evaluate devices that assess physical behavior. Exerc Sport Sci Rev 2019 Oct;47(4):206-214. [doi: 10.1249/JES.0000000000000206] [Medline: 31524786]

235. Sperlich B, Holmberg HC. Wearable, yes, but able…?: it is time for evidence-based marketing claims!. Br J Sports Med 2017 Aug;51(16):1240 [FREE Full text] [doi: 10.1136/bjsports-2016-097295] [Medline: 27986762]
236. Mühlen JM, Stang J, Lykke Skovgaard E, 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 Jul;55(14):767-779 [FREE Full text] [doi: 10.1136/bjsports-2020-103148] [Medline: 33397674]
237. McCarney R, Warner J, Iliffe S, van Haselen R, Griffin M, Fisher P. The Hawthorne Effect: a randomised, controlled trial. BMC Med Res Methodol 2007 Jul 03;7:30 [FREE Full text] [doi: 10.1186/1471-2288-7-30] [Medline: 17608932]
238. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR Mhealth Uhealth 2020 Sep 08;8(9):e18694 [FREE Full text] [doi: 10.2196/18694] [Medline: 32897239]
239. Moore CC, McCullough AK, Aguiar EJ, Ducharme SW, Tudor-Locke C. Toward harmonized treadmill-based validation of step-counting wearable technologies: a scoping review. J Phys Act Health 2020 Jul 11;17(8):840-852 [FREE Full text] [doi: 10.1123/jpah.2019-0205] [Medline: 32652514]
240. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011 Oct 18;155(8):529-536 [FREE Full text] [doi: 10.7326/0003-4819-155-8-201110180-00009] [Medline: 22007046]
241. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
242. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. BMC Med Inform Decis Mak 2007 Jun 15;7:16 [FREE Full text] [doi: 10.1186/1472-6947-7-16] [Medline: 17573961]
243. Grandner MA, Rosenberger ME. Actigraphic sleep tracking and wearables: Historical context, scientific applications and guidelines, limitations, and considerations for commercial sleep devices. In: Grandner MA, editor. Sleep and Health. Cambridge, MA, USA: Academic Press; 2019:147-157.
244. Ancoli-Israel S, Martin JL, Blackwell T, Buenaver L, Liu L, Meltzer LJ, et al. The SBSM guide to actigraphy monitoring: clinical and research applications. Behav Sleep Med 2015;13 Suppl 1:S4-38. [doi: 10.1080/15402002.2015.1046356] [Medline: 26273913]
245. Umemneku Chikere CM, Wilson K, Graziadio S, Vale L, Allen AJ. Diagnostic test evaluation methodology: a systematic review of methods employed to evaluate diagnostic tests in the absence of gold standard - an update. PLoS One 2019 Oct 11;14(10):e0223832 [FREE Full text] [doi: 10.1371/journal.pone.0223832] [Medline: 31603953]
246. Welk GJ, Bai Y, Lee JM, Godino J, Saint-Maurice PF, Carr L. Standardizing analytic methods and reporting in activity monitor validation studies. Med Sci Sports Exerc 2019 Aug;51(8):1767-1780 [FREE Full text] [doi: 10.1249/MSS.0000000000001966] [Medline: 30913159]
247. Rowe DA, Mahar MT, Raedeke TD, Lore J. Measuring physical activity in children with pedometers: reliability, reactivity, and replacement of missing data. Pediatr Exerc Sci 2004 Nov;16(4):343-354. [doi: 10.1123/pes.16.4.343]
248. Driller MW, Dunican IC. No familiarization or 'first-night effect' evident when monitoring sleep using wrist actigraphy. J Sleep Res 2021 Aug;30(4):e13246 [FREE Full text] [doi: 10.1111/jsr.13246] [Medline: 33289237]
249. Freedson P, Bowles HR, Troiano R, Haskell W. Assessment of physical activity using wearable monitors: recommendations for monitor calibration and use in the field. Med Sci Sports Exerc 2012 Jan;44(1 Suppl 1):S1-S4 [FREE Full text] [doi: 10.1249/MSS.0b013e3182399b7e] [Medline: 22157769]
250. Clemes SA, Matchett N, Wane SL. Reactivity: an issue for short-term pedometer studies? Br J Sports Med 2008 Jan;42(1):68-70. [doi: 10.1136/bjsm.2007.038521] [Medline: 18178685]
251. Staudenmayer J, Zhu W, Catellier DJ. Statistical considerations in the analysis of accelerometry-based activity monitor data. Med Sci Sports Exerc 2012 Jan;44(1 Suppl 1):S61-S67. [doi: 10.1249/MSS.0b013e3182399e0f] [Medline: 22157776]
252. Quante M, Kaplan ER, Rueschman M, Cailler M, Buxton OM, Redline S. Practical considerations in using accelerometers to assess physical activity, sedentary behavior, and sleep. Sleep Health 2015 Dec;1(4):275-284. [doi: 10.1016/j.sleh.2015.09.002] [Medline: 29073403]
253. Fairclough SJ, Noonan R, Rowlands AV, Van Hees V, Knowles Z, Boddy LM. Wear compliance and activity in children wearing wrist- and hip-mounted accelerometers. Med Sci Sports Exerc 2016 Feb;48(2):245-253. [doi: 10.1249/MSS.0000000000000771] [Medline: 26375253]
254. 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 [FREE Full text] [doi: 10.1007/s40279-017-0716-0] [Medline: 28303543]
255. 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]

256. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Yoo JY, et al. Accuracy of Fitbit devices: systematic review and narrative syntheses of quantitative data. JMIR Mhealth Uhealth 2018 Aug 09;6(8):e10527 [FREE Full text] [doi: 10.2196/10527] [Medline: 30093371]
257. Dixon PM, Saint-Maurice PF, Kim Y, Hibbing P, Bai Y, Welk GJ. A primer on the use of equivalence testing for evaluating measurement agreement. Med Sci Sports Exerc 2018 Apr;50(4):837-845 [FREE Full text] [doi: 10.1249/MSS.0000000000001481] [Medline: 29135817]
258. Choi L, Liu Z, Matthews CE, Buchowski MS. Validation of accelerometer wear and nonwear time classification algorithm. Med Sci Sports Exerc 2011 Feb;43(2):357-364 [FREE Full text] [doi: 10.1249/MSS.0b013e3181ed61a3] [Medline: 20581716]

Abbreviations

24-HAC: 24-hour activity cycle
PA: physical activity
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies-2
SB: sedentary behavior

Edited by L Buis; submitted 12.01.22; peer-reviewed by A Naseri Jahfari, K Fadahunsi; comments to author 03.03.22; revised version received 27.04.22; accepted 29.04.22; published 09.06.22.

Please cite as:
Giurgiu M, Timm I, Becker M, Schmidt S, Wunsch K, Nissen R, Davidovski D, Bussmann JBJ, Nigg CR, Reichert M, Ebner-Priemer UW, Woll A, von Haaren-Mack B
Quality Evaluation of Free-living Validation Studies for the Assessment of 24-Hour Physical Behavior in Adults via Wearables: Systematic Review
JMIR Mhealth Uhealth 2022;10(6):e36377
URL: https://mhealth.jmir.org/2022/6/e36377
doi:10.2196/36377
PMID:35679106

Original Paper

The Effect and Feasibility of mHealth-Supported Surgical Site Infection Diagnosis by Community Health Workers After Cesarean Section in Rural Rwanda: Randomized Controlled Trial

Fredrick Kateera[1*], MD, MSc, PhD; Robert Riviello[2,3*], MD, MPH; Andrea Goodman[4], MPH; Theoneste Nkurunziza[1,5], MPH, MSc; Teena Cherian[4], MPH; Laban Bikorimana[1], MPH; Jonathan Nkurunziza[1], MPH; Evrard Nahimana[1], MD, MSc; Caste Habiyakare[6], MPH; Georges Ntakiyiruta[7], MD; Alexi Matousek[2], MD; Erick Gaju[8], MSc; Magdalena Gruendl[2,5], MD, MPH; Brittany Powell[2], MD; Kristin Sonderman[2], MD, MPH; Rachel Koch[2,9], MD; Bethany Hedt-Gauthier[2,4], PhD

[1]Partners in Health, Kigali, Rwanda

[2]Center for Surgery and Public Health, Brigham and Women's Hospital, Boston, MA, United States

[3]Program in Global Surgery and Social Change, Harvard Medical School, Boston, MA, United States

[4]Department of Global Health and Social Medicine, Harvard Medical School, Boston, MA, United States

[5]Epidemiology, Department for Sport and Health Sciences, Technical University of Munich, Munich, Germany

[6]Kirehe District Hospital, Ministry of Health, Kirehe, Rwanda

[7]Ejo Heza Surgical Center, Kigali, Rwanda

[8]Rwanda Ministry of Health, Kigali, Rwanda

[9]Vanderbilt University Medical Center, Nashville, TN, United States

[*]these authors contributed equally

Corresponding Author:
Robert Riviello, MD, MPH
Center for Surgery and Public Health
Brigham and Women's Hospital
One Brigham Circle
1620 Tremont St., Suite 2-016
Boston, MA, 02120
United States
Phone: 1 6177327715
Email: rriviello@bwh.harvard.edu

Abstract

Background: The development of a surgical site infection (SSI) after cesarean section (c-section) is a significant cause of morbidity and mortality in low- and middle-income countries, including Rwanda. Rwanda relies on a robust community health worker (CHW)–led, home-based paradigm for delivering follow-up care for women after childbirth. However, this program does not currently include postoperative care for women after c-section, such as SSI screenings.

Objective: This trial assesses whether CHW's use of a mobile health (mHealth)–facilitated checklist administered in person or via phone call improved rates of return to care among women who develop an SSI following c-section at a rural Rwandan district hospital. A secondary objective was to assess the feasibility of implementing the CHW-led mHealth intervention in this rural district.

Methods: A total of 1025 women aged ≥18 years who underwent a c-section between November 2017 and September 2018 at Kirehe District Hospital were randomized into the three following postoperative care arms: (1) home visit intervention (n=335, 32.7%), (2) phone call intervention (n=334, 32.6%), and (3) standard of care (n=356, 34.7%). A CHW-led, mHealth-supported SSI diagnostic protocol was delivered in the two intervention arms, while patients in the standard of care arm were instructed to adhere to routine health center follow-up. We assessed intervention completion in each intervention arm and used logistic regression to assess the odds of returning to care.

Results: The majority of women in Arm 1 (n=295, 88.1%) and Arm 2 (n=226, 67.7%) returned to care and were assessed for an SSI at their local health clinic. There were no significant differences in the rates of returning to clinic within 30 days (P=.21), with high rates found consistently across all three arms (Arm 1: 99.7%, Arm 2: 98.4%, and Arm 3: 99.7%, respectively).

Conclusions: Home-based post–c-section follow-up is feasible in rural Africa when performed by mHealth-supported CHWs. In this study, we found no difference in return to care rates between the intervention arms and standard of care. However, given our previous study findings describing the significant patient-incurred financial burden posed by traveling to a health center, we believe this intervention has the potential to reduce this burden by limiting patient travel to the health center when an SSI is ruled out at home. Further studies are needed (1) to determine the acceptability of this intervention by CHWs and patients as a new standard of care after c-section and (2) to assess whether an app supplementing the mHealth screening checklist with image-based machine learning could improve CHW diagnostic accuracy.

Trial Registration: ClinicalTrials.gov NCT03311399; https://clinicaltrials.gov/ct2/show/NCT03311399

(JMIR Mhealth Uhealth 2022;10(6):e35155) doi:10.2196/35155

KEYWORDS

obstetric surgery; community health workers; mobile health; surgical site infections; c-section; infection; community health; Rwanda

Introduction

Rates of cesarean section (c-section) births are increasing in low- and middle-income countries (LMIC), including in sub-Saharan Africa (SSA) [1]. Increased access to timely c-section can prevent maternal and neonatal mortality, but also carries risk of perioperative complications [2]. Surgical site infections (SSIs) are a significant cause of morbidity and mortality globally, but the magnitude of the risk is significantly higher in LMIC. In SSA, post–c-section SSI rates range from 7% to 48% [3-7], in part due to geographic and infrastructural barriers that delay or prevent patients from accessing care postoperatively [8,9].

In much of SSA, networks of community health workers (CHWs) provide home-based prenatal care to pregnant women, postpartum care for women after vaginal delivery only, and follow-up care for children under the age of 5 years [10]. However, women delivering via c-section can only access follow-up care at their local health center because CHWs are not currently trained to conduct home-based postoperative follow-up or wound care. In complementary work from our team, we found that geographical and financial barriers can lead to delays in return to care after discharge [8,11]. This delayed or lack of access to care may contribute to post–c-section SSI rates, which our group reported to be 10.9% in the district where this study took place [8]. Strengthening the CHW workforce to provide SSI screening and home-based care to women who deliver via c-section could reduce barriers to care and lead to earlier detection and treatment of SSIs. However, it is not known if home-based care by CHWs is feasible or improves access to care in this context. In LMIC, previous studies have demonstrated the feasibility of phone-based surveillance of postdischarge SSI, including in women who had undergone c-section surgery [12-15]. In our study, we explored the feasibility and impact of return to care of CHW-led SSI surveillance in patient homes using a mobile health (mHealth) checklist administered via REDCap (Research Electronic Data Capture; Vanderbilt University) either in person or via phone call to facilitate remote diagnoses. This mHealth screening protocol is a battery of questions about the presence or absence of clinical findings highly associated with SSI (eg, pain, swelling, discharge, and wound gaping), which we described in previous work [16].

In this paper, we describe two CHW-led mHealth interventions to diagnose SSIs following c-section in rural Rwanda, which are (1) administering a mobile phone–based SSI screening protocol to a patient via phone call; and (2) administering the same screening protocol, carried on an electronic tablet, in person during a home visit along with collecting wound photo images on the same tablet for remote diagnosis. Here, we compare these two interventions to the standard of care via a 3-arm randomized controlled trial (ClinicalTrials.gov NCT03311399) and describe the feasibility of the two interventions in this context.

Methods

Study Setting

This study was conducted in the Eastern Province of Rwanda at Kirehe District Hospital (KDH), a 233-bed facility operated by the Rwanda Ministry of Health and supported by Partners In Health, an international NGO. KDH serves a catchment area of 364,000 people including patients from Mahama Refugee Camp, which comprises over 50,000 people [17]. In Rwanda, over 91% of women deliver in health facilities [18]. Women in labor first present to their local health center where most vaginal deliveries take place. Complex cases and cases requiring surgical intervention are transferred to the district hospital to be assessed and managed by a general practitioner, who performs the c-section procedure, if indicated. After surgery, the woman is admitted to the postoperative ward for an average of 3 days for monitoring, medication administration, and wound checks. Before leaving the hospital, she receives postdischarge instructions directing her to the health center nearest to her home for follow-up and wound dressing changes.

Study Population

This study included women aged ≥18 years who received a c-section at KDH between November 2, 2017, and September 4, 2018, and were residents of Kirehe District. Women who

developed an SSI while being an inpatient or who remained inpatient at KDH past postoperative day (POD) 10 were excluded as they were not able to participate in a POD 10 home visit. Women who resided in Mahama Refugee Camp and were therefore not covered by the CHW network were also excluded.

Preimplementation Procedures

We hired 4 study CHWs (sCHWs) using the Rwanda Ministry of Health criteria though a community-led process. The sCHWs received a 4-week training on implementation procedures, including the following: education on the Rwandan health sector; post–c-section follow-up; operating mHealth tools; best practices in wound photography; basic SSI physiopathology (including signs and symptoms of an infection); and how to examine surgical wounds and change wound dressings. Previously, we led a 7-month SSI protocol development study at KDH to identify a simple screening protocol with high accuracy to diagnose SSIs. In that study, three questions were found to have sufficient sensitivity and specificity for SSI diagnosis, which comprised the following: (1) fever since discharge from the hospital, (2) increasing pain since discharge, and (3) the presence of discolored wound discharge [19].

Intervention Implementation

This randomized study included three arms—Arm 1, where the sCHW visited a participant's home on POD 10 (SD 3 days) and administered the SSI screening protocol; Arm 2, where the sCHW called the participant and administered the SSI screening protocol over the phone on POD 10 (SD 3 days); and Arm 3, where the participant received the standard of care instructions to return to the health center for follow-up. In Arm 2, the sCHW attempted phone calls while sitting in the study office and made 3 call attempts before deeming a patient to be inaccessible. If a patient was deemed inaccessible, they were included in the evaluation of the feasibility of the intervention and classified as "not successfully assessed for SSI." However, these individuals were not included in the evaluation of the presence of an SSI. If a participant in Arm 1 or 2 was suspected of having an SSI by the responses to the screening protocol, the sCHW was prompted to refer her to a nearby health center for additional medical care.

Enrollment, Randomization, Follow-up, and Data Collection

All data collected were entered and stored using REDCap (v8.10.20), a secure web application certified for medical research studies [20]. The study staff enrolled and randomized eligible participants at discharge, independent of any patient factors, to one of the three study arms. The study staff prepared study packets in sealed envelopes numbered consecutively. REDCap was then used to randomly generate arm assignments to each packet using simple randomization in a 1:1:1 ratio [16]. All consenting participants' demographic and socioeconomic data were collected using a self-reported questionnaire administered by a trained study data collector. In addition, the study staff extracted clinical data from the patients' medical files. Upon discharge, the patients received a packet with arm-specific follow-up and general discharge instructions.

Each health center in the catchment area and KDH had a study-specific patient registry to document return to care. Study staff entered the following details into REDCap: return to care status, SSI diagnosis (by nurse), treatment received, hospitalization, patient referral, and need for surgical procedures, if any. Data collectors contacted each woman on POD 30 to validate what was captured in the registry and to ensure that no follow-up visits were missed.

Statistical Analysis

Data were analyzed using Stata (14.0 version, Stata Corp) statistical software. We characterized study participant demographics and clinical characteristics using descriptive statistics. For the primary outcome, a patient was classified as having returned to care if the return visit was documented in the health center registry or if the patient reported returning to the health center during the POD 30 follow-up call. In this primary analysis, we excluded anyone without information on return to care by POD 30 from analyses. Feasibility assessments across Arms 1 and 2 were reported as the percentage of visits where that specific task was completed. We used a Fisher exact test at α=.05 significance level to assess the association between patients' return to care and interventions implemented in Arms 1 and 2 as compared to Arm 3 (standard of care). We used a logistic regression model to assess the impact of study interventions on return to care, controlling for potential confounders that were unbalanced at baseline. In this primary analysis, we excluded anyone without information on return to care by POD 30 from analyses. We used chi-squared tests to assess for differences in having information about return to care by study arm and by patient demographics. We also conducted a sensitivity analysis, whereby any individual missing information on return to care was presumed to have not returned to care.

Power

The estimated sample size was 364 patients per arm, for a total of 1092 patients. We anticipated an SSI rate of 15%, which would result in 55 SSIs per arm. Assuming an 80% return to care rate in the two intervention arms and a 40% return to care rate in the standard of care arm, we would have 81% power to detect a difference between the two intervention arms as compared to the routine care arm. The trial was halted when 1166 patients were enrolled (in excess of the targeted sample size of 1092).

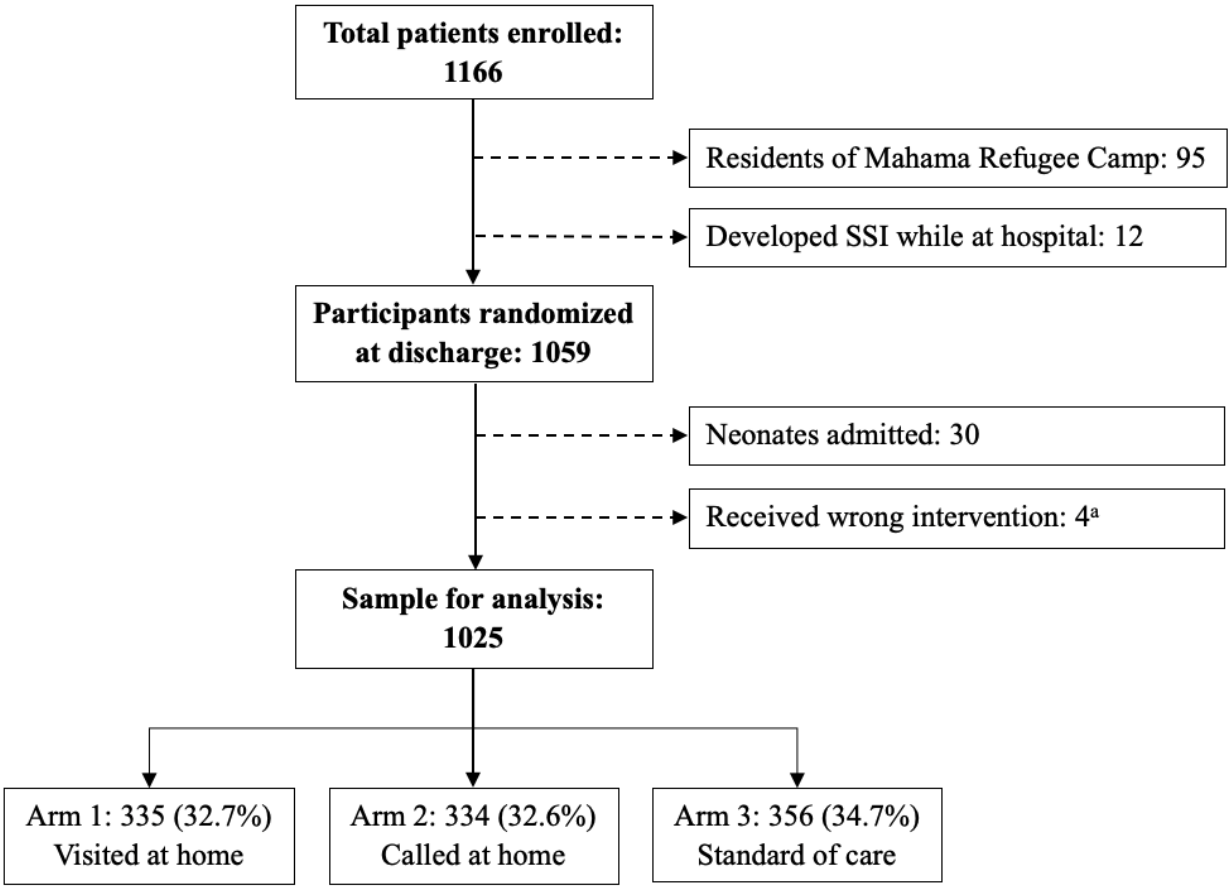
Ethical Considerations

Eligible women gave informed consent prior to participation. The study team members provided information in Kinyarwanda, including details of the three study arms and the right to withdraw from the study or refrain from giving information at any stage. Deidentified data were collected and managed using REDCap. This study was approved by the Rwanda National Ethics Committee (848/RNEC/2016) and Partners Human Research Committee (2016P001943/MGH). Seven months into the study, a Data and Safety Monitoring Board reviewed the study participants' safety, data quality, and midterm outcomes, and deemed it appropriate to continue to study completion.

Results

In total, 1166 women were enrolled, of which 107 (9.1%) were excluded—95 residents of Mahama Refugee Camp and 12 patients who developed an SSI while at the hospital. Of the enrolled patients who were randomized at discharge, 34 participants were removed from analysis—30 participants who remained in hospital after discharge to attend to their admitted neonates and 4 participants who were assigned to one arm but inadvertently received the follow-up of another arm. Of the remaining 1025 women, 335 (32.7%) were randomized to Arm 1, 334 (32.6%) to Arm 2, and 356 (34.7%) to Arm 3 (Figure 1).

Figure 1. Flow chart of patient randomization into two treatment arms and one standard of care arm. a: one patient from Arm 1 received a phone call instead of an in-person visit, and 3 patients from Arm 3 received the in-person intervention despite being randomized into the control group; these 4 patients were excluded from analysis.



There were no significant differences between the three groups for most demographic variables (Table 1), including age (*P*=.29), marital status (*P*=.2), occupation (*P*=.496), or type of insurance (*P*=.15). The only statically significant differences found were regarding education and income. Women in Arm 3 were more likely to report having only a primary education (*P*=.006). Women in Arm 2 were significantly more likely to report higher income (*P*=.03). There was no significant difference between the three groups in terms of their access to health care, measured by the cost of transportation from the woman's home to the nearest health center (*P*=.93) and the travel time from a woman's home to the nearest health center (*P*=.25; Table 1).

Table 1. Demographic characteristics of study participants by study arm (n=1025).

Characteristics	Total, n (%)	Arm 1: home visit (n=335), n (%)	Arm 2: phone call (n=334), n (%)	Arm 3: standard of care (n=356), n (%)	*P* value
Age (years)					.28
18-21	174 (17.0)	52 (15.5)	50 (15.0)	72 (20.2)	
22-30	549 (53.6)	182 (54.3)	178 (53.3)	189 (53.1)	
>30	302 (29.5)	101 (30.2)	106 (31.7)	95 (26.7)	
Education					.006
No education	90 (8.8)	29 (8.7)	39 (11.7)	22 (6.2)	
Primary education	696 (67.9)	225 (67.2)	205 (61.4)	266 (74.7)	
Secondary education	214 (20.9)	75 (22.4)	82 (24.6)	57 (16.0)	
University education	25 (2.4)	6 (1.8)	8 (2.4)	11 (3.1)	
Marital status					.2
Single	95 (9.3)	28 (8.4)	26 (7.8)	41 (11.5)	
Married	446 (43.5)	156 (46.6)	148 (44.3)	142 (39.9)	
Living with a partner	480 (46.8)	150 (44.8)	160 (47.9)	170 (47.8)	
Separated (divorced or widowed)	4 (0.4)	1 (0.3)	0 (0)	3 (0.8)	
Occupation					.496
Student	10 (1.0)	2 (0.6)	5 (1.5)	3 (0.8)	
Farmer	874 (85.3)	289 (86.3)	280 (83.8)	305 (85.7)	
Employed	38 (3.7)	8 (2.4)	15 (4.5)	15 (4.2)	
Self-employed	69 (6.7)	28 (8.4)	21 (6.3)	20 (5.6)	
Housewife	34 (3.3)	8 (2.4)	13 (3.9)	13 (3.7)	
Income[a] (US $)					.03
>33.70	854 (83.3)	290 (86.6)	264 (79.0)	300 (84.3)	
<33.70	171 (15.7)	45 (13.4)	70 (21.0)	56 (15.7)	
Type of insurance					.15
No insurance	24 (2.3)	5 (1.5)	10 (3.0)	9 (2.5)	
Community-based insurance	941 (91.8)	316 (94.3)	297 (88.9)	328 (92.1)	
Private insurance	60 (5.9)	14 (4.2)	27 (8.1)	19 (5.3)	
Cost of transportation from home to health center[a] (US $; n=969)					.93
≤1.12	586 (60.5)	192 (60.0)	192 (61.3)	202 (60.1)	
>1.12	383 (39.5)	128 (40.0)	121 (38.7)	134 (39.9)	
Time from home to health center (n=964)					.25
≤1 hour	881 (91.4)	288 (90.0)	281 (90.7)	312 (93.4)	
>1 hour	83 (8.6)	32 (10.0)	29 (9.4)	22 (6.7)	

[a]Calculated using an exchange rate of US $1 to 890 Rwandan Francs.

Of the 335 women in Arm 1, 295 (88.1%) were successfully visited in their homes and had the full SSI assessment completed by the sCHW (Table 2). The primary reasons for noncompletion in Arm 1 were prolonged hospitalization of either mother or baby or an inability to contact the mother to confirm the home visit appointment. Of the 334 women in Arm 2, 67.7% (n=226) were successfully called and assessed over the phone by a sCHW for an SSI. The primary reasons for noncompletion in Arm 2 were as follows: lack of mobile phone ownership, poor network coverage, or the phone belonging to another person (eg, husband or neighbor). Women in Arm 1 had slightly higher rates of reporting SSI symptoms as compared to women in Arm 2 (Table 3). As Arm 3 was the standard of care arm, there was no attempt to contact patients either via phone call or home visit.

Table 2. Feasibility of community health worker intervention arms.

Interventions and call attempts	Values, n (%)
Home visit intervention (n=335)	
Number of patients who were visited and assessed for SSI[a] by CHW[b] at patient's home	295 (88.1)
Home visits attempted by study CHW	295 (88.1)
Completion of steps to conduct home visit	
Study CHW was able to find local CHW in patient's village	281 (95.3)
Study CHW was able to locate patient's home	295 (100)
Study CHW was allowed into patient's home	295 (100)
Patient was at home when study CHW arrived	287 (97.3)
Patient allowed study CHW to ask SSI screening questions	295 (100)
Patient allowed study CHW to physically examine her	295 (100)
Phone call intervention (n=334)	
Phone call to patient attempted by study CHW	319 (95.5)
Number of patients who were called and assessed for SSI by CHW over phone	226 (67.7)
Phone call attempt #1	
Phone number went through, or phone rang (n=319)	268 (84)
Phone call resulted in talking with the patient (n=268)	167 (62.3)
Outcomes of talking with patient (n=167)	
Patient answered SSI screening questions at time of call	163 (97.6)
Patient was busy	3 (1.8)
Patient did not respond, reason not recorded	1 (0.6)
Reason for not talking with patient (n=101)	
Wrong number	7 (6.9)
Patient did not pick up the phone	6 (5.9)
Another person picked up the phone, patient was not available	87 (86.1)
Not reported	1 (1.0)
Patients requiring a second attempt (n=319)	156 (48.9)
Phone call attempt #2	
Number of patients who were called a second time (n=156)	133 (85.3)
Phone number went through, or phone rang (n=133)	89 (66.9)
Phone call resulted in talking with the patient (n=89)	51 (57.3)
Outcomes of talking with patient (n=51)	
Patient answered SSI screening questions at time of call	50 (98)
Patient did not respond, reason not recorded	1 (2)
Reason for not talking with patient (n=38)	
Patient did not pick up the phone	4 (11)
Another person picked up the phone, patient was not available	33 (89)
Not reported	1 (3)
Patients requiring a third attempt (n=156)	106 (67.9)
Phone call attempt #3	
Number of patients who were called a second time (n=106)	83 (78.3)
Phone number went through/phone rang (n=83)	36 (43)
Phone call resulted in talking with the patient (n=36)	13 (36)

Interventions and call attempts	Values, n (%)
Outcomes of talking with patient (n=13)	
Patient answered SSI screening questions at time of call	13 (100)
Reason for not talking with patient (n=23)	
Wrong number	1 (4)
Patient did not pick up the phone	2 (9)
Another person picked up the phone, patient was not available	20 (87)

[a]SSI: surgical site infection.

[b]CHW: community health worker.

Table 3. CHW[a] screening results by study arm (n=523)[b].

Responses to CHW screening	Arm 1: home visit (n=295), n (%)	Arm 2: phone call (n=228), n (%)
Fever since discharge (n=522)	35[c] (11.9)	22 (9.7)
Pain since discharge	51 (17.3)	32 (14.0)
Discolored drainage since discharge	46 (15.6)	20 (8.8)
CHW suspected wound infection	52 (17.6)	29 (12.7)
CHW advised patient to return to care (n=482)[d]	56[e] (20.3)	30[f] (14.6)

[a]CHW: community health worker.

[b]Arm 3 not included because no CHW screenings occurred in the standard of care arm.

[c]Missing data for 1 patient, n=294.

[d]Among those for whom CHW suspected wound infection in the home visit arm, 1 patient was not advised to return to care.

[e]Missing data for 19 patients, n=276.

[f]Missing data for 22 patients, n=206.

We had information on return to care for 896/1025 (87.4%) women, as described in Table 4. Women in Arm 2 were marginally, but nonsignificantly, more likely to have this information recorded (*P*=.06). There were no differences in having this documented among key demographics; though women with higher monthly incomes were more likely to have information on return to care recorded (*P*=.03). In the primary analyses, there was no difference in care-seeking behavior between the three arms. Women across all three arms had high rates of returning to clinic by POD 30 (99.7% in Arm 1, 98.4% in Arm 2, and 99.7% in Arm 3), with no significant statistical difference between them (*P*=.21 crude; *P*=.19 adjusted). Reasons for returning to care were not significantly different between the groups, with similar percentages of women returning for either routine wound care (n=253, 89.4% in Arm 1; n=264, 88.6% in Arm 2; and n=278, 90.3% in Arm 3; *P*=.08) or for a specific concern related to their c-section (n=30, 10.6% in Arm 1; n=34, 11.4% in Arm 2; and n=30, 9.7% in Arm 3; *P*=.08). There were similar rates of nurse-diagnosed SSIs in each group (n=33, 11.9% in Arm 1; n=34, 11.6% in Arm 2; and n=28, 9.3% in Arm 3; *P*=.54). In the sensitivity analysis, difference in return to care rates by study arm remained insignificant (*P*=.19 crude; *P*=.26 adjusted).

Table 4. Return to care behavior by 30-day post–c-section[a] by study arm (n=896).

Outcomes	Total, n (%)	Arm 1: home visit, n (%)	Arm 2: phone call, n (%)	Arm 3: standard of care, n (%)	*P* value
Patients randomized	1025 (100)	335 (32.7)	334 (32.6)	356 (34.7)	N/A[b]
Patients[c] with 30-day follow-up data	896 (87.4)	284 (84.8)	303 (90.7)	309 (68.8)	N/A
Source of 30-day follow-up data[d]					
Phone call with patient	555 (61.9)	180 (63.4)	188 (62.1)	187 (60.5)	.77
Health center registry	635 (70.9)	194[e] (68.6)	215 (71.0)	226 (73.1)	.47
District hospital medical records	18 (2.0)	11 (3.9)	2 (0.7)	5 (1.6)	.02
Patients[f] who returned to care (n=896)	889 (99.2)	283 (99.7)	298 (98.4)	308 (99.7)	.21[g]
Among those with 30-day follow-up data (n=889)					
Reason for returning to care					.8
Routine wound care (wound check and removal of stitches)	795 (89.4)	253 (89.4)	264 (88.6)	278 (90.3)	
Concern related to c-section (fever, pain, and concern about wound)	94 (10.6)	30 (10.6)	34 (11.4)	30 (9.7)	
Patient returned to care with nurse-diagnosed SSI[h] (n=871)	95 (10.7)	33[i] (11.9)	34[j] (11.6)	28[k] (9.3)	.54

[a]c-section: cesarean section.

[b]N/A: not applicable.

[c]Those who were randomized.

[d]Information could have been collected from more than one source.

[e]Missing data for 1 patient (n=283).

[f]Those with 30-day follow-up data.

[g]*P*=.19 from likelihood ratio test (from logistic regression models controlling for education and income).

[h]SSI: surgical site infection.

[i]Missing data for 5 patients (n=278).

[j]Missing data for 6 patients (n=292).

[k]Missing data for 7 patients (n=301).

Discussion

Principal Findings

Surprisingly, nearly all patients in our study returned to care at least once by POD 30, with no significant difference in follow-up between arms. This contrasts with the findings in the Central African Republic, where a study reported that only 25% of surgical patients returned for a POD 30 follow-up visit [21]. A possible reason for this is that Rwanda, a small country with a strong functional decentralized public health system [22], offers greater access to follow-up care.

In this study, we observed that home-based follow-up care of participants allows the sCHW to enter the women's homes, physically examine them, and take a photo of their wound. As close to 90% of participants in Arm 1 successfully visited and were assessed for SSI, we found that home visits are a feasible way to conduct post–c-section care. In rural Rwanda and many other low-resource settings, CHWs already provide in-home screening for child health [10], maternal health [23], and HIV care [24] referrals. The high feasibility of in-home screening could be due to the familiarity women have with the CHW system and how they value support from CHWs [24,25].

On the other hand, SSI screening by phone excluded close to 30% of women. Other studies in Tanzania [12] and Sudan [26] also demonstrated gaps in using telephone calls for postdischarge surveillance of SSIs. Currently, only 54% of households in Rwanda own a mobile phone [18]. Telephone-based interventions may be more feasible as phone access and network coverage expand. Two recent systematic review articles assessing the use of smartphones to identify SSI found that there are few articles in the literature, the majority are in high income settings, and they require smartphone ownership by patients. We have not found any other experience of CWH home follow-up for SSI identification and care [27,28]. Currently in the rural Rwandan setting, in-person sCHW visitation at the patient's home provides greater follow-up coverage than phone calls alone.

Despite the null results in the difference between rates of return to care, this study has important implications linked to our understanding of the financial risks associated with health care seeking in this population. Undergoing a c-section is a financial

burden for women in rural Rwanda [29]. This is true even for women covered by community-based health insurance. Our group has previously reported that the median out-of-pocket cost of transport for a single visit to the health center for women who received a c-section at KDH is up to 10% of their monthly income and that those who spent more money had increased risk of SSI [30]. Transportation cost was also self-reported by patients from Arm 3 of the study to be a barrier to health care seeking in the postoperative period [11]. These costs of transport were uniform across the arms of the study, as all three arms had equal rates of return to clinic.

In this setting, approximately 90% of patients do not develop an SSI. Treatment of an SSI is a principal reason for return to health center for care that cannot be provided by a CHW. Thus, accurate home rule-out of a post–c-section SSI can eliminate the need for 90% of women undergoing c-section to make the return journey and incur the out-of-pocket expense of transport to the health center. Given the financial burden of transport to the health center and the feasibility of in-person CHW SSI screening, leveraging the existing CHW system in Rwanda to bring postcesarean care to women's homes could reduce both financial barriers to care and medical impoverishment. Further analysis is needed to determine the effect that home-based surgical wound monitoring would have on reducing unnecessary visits to the health center, health system cost savings, and workload on clinicians, though promising results have been shown in other settings [31,32]. This study's findings will also be used to develop a supporting app to facilitate in-person SSI screening by CHWs.

Limitations

This study had several limitations. Health center data may not have been consistent in quality due to variations in study patient tracking and data collection processes across sites. Targeted interventions, including calling health centers weekly with a list of expected enrolled patients and monthly in-person audits of each health center's registry, were implemented to improve patient tracking. We also called all patients on POD 30 regarding their follow-up activities and SSI diagnoses. There was 100% agreement between data from registries and phone calls for the 228 patients from Arms 1 and 2 [19]. Secondly, we could not accurately or consistently capture dates of health center visits. As a result, we do not know when within-POD-30 women returned to care and whether there were differences between the intervention arms and the control arm. Additional research is needed to assess how CHW interventions affect the timeliness of return to care for post–c-section SSI evaluation.

Conclusions

We did not observe a difference in the rate of return to the health center between women who were visited at home, who called at home, and who asked to continue with standard of care visits. In fact, women in all groups demonstrated high levels of health seeking behavior. However, our study found that home-based post–c-section follow-up by CHWs facilitated by an mHealth app to identify and refer SSIs is feasible. Our previous studies have shown that health center visits can pose a significant financial burden on women following c-section. Therefore, use of home visits for postoperative care could greatly reduce the nonmedical costs related to transport for routine follow-up for women who do not develop SSIs. Home, mHealth-enhanced, visits were also found to be more effective than phone-based follow-up for connecting CHWs with patients. Thus, home visits have the potential to greatly reduce the patient's economic burden of post–c-section care. Future studies to understand the acceptability of CHW home visits for patients and health care workers are needed before this can be adopted as a standard care protocol.

Acknowledgments

This work was supported by National Institutes of Health grant 1R21EB022369.

Authors' Contributions

FK, RR, and BHG are the primary investigators for the National Institute of Health grant that provided funding for this research. FK, RR, and BHG led the design and planning of this study, supervised the conduct and data analysis, and were involved in all steps of manuscript drafting and revising. AG led the drafting and revising of the manuscript. TN, TC, LB, EN, CH, GN, AM, EG, LG, BP, KS, RK, and JN were involved in study design, planning, conduct, and data analysis, and were involved in manuscript drafting and editing.

Conflicts of Interest

None to declare.

Multimedia Appendix 1
CONSORT-EHEALTH (V 1.6.1) checklist.
[PDF File (Adobe PDF File), 2873 KB - mhealth_v10i6e35155_app1.pdf]

References

1. Harrison MS, Goldenberg RL. Cesarean section in sub-Saharan Africa. Matern Health Neonatol Perinatol 2016;2:6 [FREE Full text] [doi: 10.1186/s40748-016-0033-x] [Medline: 27398224]

2. Statement on caesarean section rates. World Health Organization. 2015. URL: https://www.who.int/reproductivehealth/publications/maternal_perinatal_health/cs-statement/en [accessed 2020-03-01]
3. Mivumbi VN, Little SE, Rulisa S, Greenberg JA. Prophylactic ampicillin versus cefazolin for the prevention of post-cesarean infectious morbidity in Rwanda. Int J Gynaecol Obstet 2014 Mar;124(3):244-247. [doi: 10.1016/j.ijgo.2013.09.017] [Medline: 24380609]
4. De Nardo P, Gentilotti E, Nguhuni B, Vairo F, Chaula Z, Nicastri E, et al. Post-caesarean section surgical site infections at a Tanzanian tertiary hospital: a prospective observational study. J Hosp Infect 2016 Aug;93(4):355-359. [doi: 10.1016/j.jhin.2016.02.021] [Medline: 27125664]
5. Mpogoro FJ, Mshana SE, Mirambo MM, Kidenya BR, Gumodoka B, Imirzalioglu C. Incidence and predictors of surgical site infections following caesarean sections at Bugando Medical Centre, Mwanza, Tanzania. Antimicrob Resist Infect Control 2014;3:25 [FREE Full text] [doi: 10.1186/2047-2994-3-25] [Medline: 25126415]
6. Chu K, Maine R, Trelles M. Cesarean section surgical site infections in sub-Saharan Africa: a multi-country study from Medecins Sans Frontieres. World J Surg 2015 Feb;39(2):350-355 [FREE Full text] [doi: 10.1007/s00268-014-2840-4] [Medline: 25358418]
7. Kaboré B, Soudouem G, Seck I, Millogo T, Evariste Yaméogo WM, Kouanda S. A case-control study of risk factors for surgical site infection after cesarean delivery in eastern Burkina Faso. Int J Gynaecol Obstet 2016 Nov;135 Suppl 1:S107-S110. [doi: 10.1016/j.ijgo.2016.08.018] [Medline: 27836076]
8. Grimes CE, Bowman KG, Dodgion CM, Lavy CBD. Systematic review of barriers to surgical care in low-income and middle-income countries. World J Surg 2011 May 1;35(5):941-950. [doi: 10.1007/s00268-011-1010-1] [Medline: 21360305]
9. Hsia RY, Mbembati NA, Macfarlane S, Kruk ME. Access to emergency and surgical care in sub-Saharan Africa: the infrastructure gap. Health Policy Plan 2012 May;27(3):234-244. [doi: 10.1093/heapol/czr023] [Medline: 21441566]
10. Mugeni C, Levine AC, Munyaneza RM, Mulindahabi E, Cockrell HC, Glavis-Bloom J, et al. Nationwide implementation of integrated community case management of childhood illness in Rwanda. Glob Health Sci Pract 2014 Aug;2(3):328-341 [FREE Full text] [doi: 10.9745/GHSP-D-14-00080] [Medline: 25276592]
11. Powell B, Nkurunziza T, Kateera F. Synthesizing postpartum care pathways, facilitators, and barriers for women after cesarean section: a qualitative analysis from rural Rwanda. J Glob Health Rep 2021 Jun 18;5:e2021043. [doi: 10.29392/001c.23615]
12. Nguhuni B, De Nardo P, Gentilotti E, Chaula Z, Damian C, Mencarini P, et al. Reliability and validity of using telephone calls for post-discharge surveillance of surgical site infection following caesarean section at a tertiary hospital in Tanzania. Antimicrob Resist Infect Control 2017;6:43 [FREE Full text] [doi: 10.1186/s13756-017-0205-0] [Medline: 28503302]
13. Aiken AM, Wanyoro AK, Mwangi J, Mulingwa P, Wanjohi J, Njoroge J, et al. Evaluation of surveillance for surgical site infections in Thika Hospital, Kenya. J Hosp Infect 2013 Feb;83(2):140-145 [FREE Full text] [doi: 10.1016/j.jhin.2012.11.003] [Medline: 23332563]
14. Pathak A, Sharma S, Sharma M, Mahadik VK, Lundborg CS. Feasibility of a Mobile Phone-Based Surveillance for Surgical Site Infections in Rural India. Telemed J E Health 2015 Nov;21(11):946-949 [FREE Full text] [doi: 10.1089/tmj.2014.0199] [Medline: 25748641]
15. Sandberg CEJ, Knight SR, Qureshi AU, Pathak S. Using Telemedicine to Diagnose Surgical Site Infections in Low- and Middle-Income Countries: Systematic Review. JMIR Mhealth Uhealth 2019 Aug 19;7(8):e13309 [FREE Full text] [doi: 10.2196/13309] [Medline: 31429414]
16. Sonderman KA, Nkurunziza T, Kateera F, Gruendl M, Koch R, Gaju E, et al. Using mobile health technology and community health workers to identify and refer caesarean-related surgical site infections in rural Rwanda: a randomised controlled trial protocol. BMJ Open 2018 May 08;8(5):e022214 [FREE Full text] [doi: 10.1136/bmjopen-2018-022214] [Medline: 29739786]
17. Mahama: Now a Safe Haven for over 50,000 Burundian refugees. The UN Refugee Agency. 2016. URL: https://www.unhcr.org/rw/559-559.html [accessed 2020-03-01]
18. Statistical yearbook. National Institute of Statistics of Rwanda (NISR). 2016. URL: http://www.statistics.gov.rw/publication/statistical-yearbook-2016 [accessed 2020-08-28]
19. Cherian T, Hedt-Gauthier B, Nkurunziza T, Sonderman K, Gruendl MA, Nihiwacu E, et al. Diagnosing Post-Cesarean Surgical Site Infections in Rural Rwanda: Development, Validation, and Field Testing of a Screening Algorithm for Use by Community Health Workers. Surg Infect (Larchmt) 2020 Sep;21(7):613-620 [FREE Full text] [doi: 10.1089/sur.2020.062] [Medline: 32423365]
20. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
21. Bercion R, Gaudeuille A, Mapouka PA, Behounde T, Guetahoun Y. [Surgical site infection survey in the orthopaedic surgery department of the "Hôpital communautaire de Bangui," Central African Republic]. Bull Soc Pathol Exot 2007 Aug;100(3):197-200. [Medline: 17824315]
22. Manzi A, Munyaneza F, Mujawase F, Banamwana L, Sayinzoga F, Thomson DR, et al. Assessing predictors of delayed antenatal care visits in Rwanda: a secondary analysis of Rwanda demographic and health survey 2010. BMC Pregnancy Childbirth 2014 Aug 28;14:290 [FREE Full text] [doi: 10.1186/1471-2393-14-290] [Medline: 25163525]

23. Dao B, Ngabo F, Zoungrana J, Rawlins B, Mukarugwiro B, Musoni P, et al. Increasing Access to Prevention of Postpartum Hemorrhage Interventions for Births in Health Facilities and at Home in Four Districts of Rwanda. Afr J Reprod Health 2015 Dec;19(4):58-67. [Medline: 27337854]
24. Mwai GW, Mburu G, Torpey K, Frost P, Ford N, Seeley J. Role and outcomes of community health workers in HIV care in sub-Saharan Africa: a systematic review. J Int AIDS Soc 2013 Sep 10;16:18586 [FREE Full text] [doi: 10.7448/IAS.16.1.18586] [Medline: 24029015]
25. Condo J, Mugeni C, Naughton B, Hall K, Tuazon MA, Omwega A, et al. Rwanda's evolving community health worker system: a qualitative assessment of client and provider perspectives. Hum Resour Health 2014 Dec 13;12:71 [FREE Full text] [doi: 10.1186/1478-4491-12-71] [Medline: 25495237]
26. Tuyisenge G, Crooks VA, Berry NS. Facilitating equitable community-level access to maternal health services: exploring the experiences of Rwanda's community health workers. Int J Equity Health 2019 Nov 26;18(1):181 [FREE Full text] [doi: 10.1186/s12939-019-1065-4] [Medline: 31771605]
27. Elbur AI, Ma Y, Elsayed ASA, Abdel-Rahman ME. Post-discharge surveillance of wound infections by telephone calls method in a Sudanese Teaching Hospital. J Infect Public Health 2013 Oct;6(5):339-346 [FREE Full text] [doi: 10.1016/j.jiph.2013.04.005] [Medline: 23999331]
28. Baniasadi T, Ghazisaeedi M, Hassaniazad M, Niakan Kalhori SR, Shahi M, Esmaeili M. Surgical Patients Follow-Up by Smartphone-Based Applications: A Systematic Literature Review. Stud Health Technol Inform 2020 Jun 23;271:85-92. [doi: 10.3233/SHTI200079] [Medline: 32578546]
29. Ng HJH, Huang D, Rajaratnam V. Diagnosing surgical site infections using telemedicine: A Systematic Review. Surgeon 2021 Jun 19:Online ahead of print. [doi: 10.1016/j.surge.2021.05.004] [Medline: 34158250]
30. Nkurunziza T, Kateera F, Sonderman K, Gruendl M, Nihiwacu E, Ramadhan B, et al. Prevalence and predictors of surgical-site infection after caesarean section at a rural district hospital in Rwanda. Br J Surg 2019 Jan;106(2):e121-e128 [FREE Full text] [doi: 10.1002/bjs.11060] [Medline: 30620071]
31. Koch R, Nkurunziza T, Irasubiza HL, Shrime M, Hedt-Gauthier B, Kateera F. Financial risk protection in cesarean section patients at a rural district hospital in Rwanda. East Cent Afr J Surg 2018;23(3):152-153.
32. Evans HL, Lober WB. A Pilot Use of Patient-Generated Wound Data to Improve Postdischarge Surgical Site Infection Monitoring. JAMA Surg 2017 Jun 01;152(6):595-596. [doi: 10.1001/jamasurg.2017.0568] [Medline: 28423162]

Abbreviations

c-section: cesarean section
CHW: community health worker
KDH: Kirehe District Hospital
LMIC: low- and middle-income countries
mHealth: mobile health
POD: postoperative day
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture
sCHW: study community health worker
SSA: sub-Saharan Africa
SSI: surgical site infection

Edited by L Buis; submitted 02.12.21; peer-reviewed by A Joseph, H Ayatollahi; comments to author 02.02.22; revised version received 30.03.22; accepted 04.05.22; published 08.06.22.

Please cite as:
Kateera F, Riviello R, Goodman A, Nkurunziza T, Cherian T, Bikorimana L, Nkurunziza J, Nahimana E, Habiyakare C, Ntakiyiruta G, Matousek A, Gaju E, Gruendl M, Powell B, Sonderman K, Koch R, Hedt-Gauthier B
The Effect and Feasibility of mHealth-Supported Surgical Site Infection Diagnosis by Community Health Workers After Cesarean Section in Rural Rwanda: Randomized Controlled Trial
JMIR Mhealth Uhealth 2022;10(6):e35155
URL: https://mhealth.jmir.org/2022/6/e35155
doi:10.2196/35155
PMID:35675108

Original Paper

A Stakeholder-Centered mHealth Implementation Inquiry Within the Digital Health Innovation Ecosystem in South Africa: MomConnect as a Demonstration Case

Idon-Nkhenso Sibuyi[1], DTech; Retha de la Harpe[1], DTech; Peter Nyasulu[2], PhD

[1]Faculty of Informatics and Design, Cape Peninsula University of Technology, Cape Town, South Africa

[2]Division of Epidemiology & Biostatistics, Faculty of Medicine & Health Sciences, Stellenbosch University, Cape Town, South Africa

Corresponding Author:
Idon-Nkhenso Sibuyi, DTech
Faculty of Informatics and Design
Cape Peninsula University of Technology
PO Box 652
Cape Town, 8000
South Africa
Phone: 27 833389132
Email: insibuyioptometrist@gmail.com

Abstract

Background: The internet is a useful web-based multimedia platform for accessing and disseminating information unconstrained by time, distance, and place. To the health care sector's benefit, the advent and proliferation of mobile devices have provided an opportunity for interventions that combine asynchronous technology-aided health services to improve the lives of the less privileged and marginalized people and their communities, particularly in developing societies.

Objective: This study aimed to report on the perspectives of the different stakeholders involved in the study and to review an existing government mobile health (mHealth) program. It forms part of a study to design a re-engineered strategy based on the best demonstrated practices (considerations and methods) and learned experiences from the perspectives of multiple stakeholders within the digital health innovation ecosystem in South Africa.

Methods: This study used an ethnographic approach involving document review, stakeholder mapping, semistructured individual interviews, focus group discussions, and participant observations to explore, describe, and analyze the perspectives of its heterogeneous participant categories representing purposively sampled but different constituencies.

Results: Overall, 80 participants were involved in the study, in addition to the 6 meetings the researcher attended with members of a government-appointed task team. In addition, 46 archived records and reports were consulted and reviewed as part of gathering data relating to the government's MomConnect project. Among the consulted stakeholders, there was general consensus that the existing government-sponsored MomConnect program should be implemented beyond mere piloting, to *as best as possible* capacity within the available resources and time. It was further intimated that the scalability and sustainability of mHealth services as part of an innovative digital health ecosystem was hamstrung by challenges that included stakeholder mismanagement, impact assessment inadequacies, management of data, lack of effective leadership and political support, inappropriate technology choices, eHealth and mHealth funding, integration of mHealth to existing health programs in tandem with Goal 3 of the Sustainable Development Goals, integration of lessons learned from other mHealth initiatives to avoid resource wastage and duplication of efforts, proactive evaluation of both mHealth and eHealth strategies, and change management and developing human resources for eHealth.

Conclusions: This study has only laid a foundation for the re-engineering of mHealth services within the digital health innovation ecosystem. This study articulated the need for stakeholder collaboration, such as continuous engagement among academics, technologists, and mHealth fieldwork professionals. Such compelling collaboration is accentuated more by the South African realities of the best practices in the fieldwork, which may not necessarily be documented in peer-reviewed or systematic research documents from which South African professionals, research experts, and practitioners could learn. Further research is needed for the retrospective analysis of mHealth initiatives and forecasting of the sustainability of current and future mHealth initiatives in South Africa.

(JMIR Mhealth Uhealth 2022;10(6):e18188) doi:10.2196/18188

KEYWORDS

MomConnect; mHealth; patient-facing eHealth; digital health innovation ecosystem; practitioner-researcher; stakeholder-centered design; re-engineering in health services; sustainable development goals; principles of digital development; global digital health index; strong structuration theory

Introduction

Background

The term *digital health innovation ecosystem* (DHIE) entails the systematization of digital health networks, environments, and communities (stakeholders) relying on information and communication technology (ICT) to connect and relate with each other for the purpose of improving health services. Furthermore, it empowers patients to manage their well-being and that of their families and communities [1]. The process of developing and designing a health service requires an in-depth understanding and knowledge of the ecosystem, as well as holistic consideration from different stakeholder perspectives [2]. Accordingly, the design process is intended for the use and distribution of a (health) product at any moment and at any of its location points to people during the product's lifetime, to the advantage of those whose needs are not appropriately addressed. In the context of this study, the centralization of the DHIE underpins the researcher's effort to explore and identify the fundamental tenets in the design, application, and implementation of a digitalized health product with respect to its entire value chain of networks, environments, and communities [3]. The networks in the ICT-dominated ecosystem or environment are constituted by health care stakeholders, institutions, and devices. Such an environment is characterized principally by the successful implementation of interactive digital best demonstrated practices and solutions [4]. In this study, the DHIE is pivotal, as it is the term or concept or phenomenon around which the implementation efficacy (or otherwise) of the MomConnect project (as a demonstration case) could be determined as an initiative of the National Department of Health (NDOH) in its efforts to introduce and implement a national rollout of the digital health strategy for South Africa [5].

Given the aforementioned context of the DHIE, this study focuses on an important topic pertaining to the implementation of the MomConnect project, which has had a nationwide rollout in South Africa. It is not a small-scale pilot project limited to several facilities or a single district but has been implemented nationwide and emphatically referred to since 2014 by previous Ministers of Health and is currently in their annual health budget speeches as an indication of the government's irreversible trajectory toward the digitization of health services as a means to broaden access [5,6]. Therefore, this research on sustainability issues and the factors affecting its success is critical.

Technology-assisted health care service delivery is regarded as the new frontier of innovations in health care, facilitated by improvements in ICTs and the internet's asynchronous interconnectivity [4]. Mobile health (mHealth) technology—part of eHealth—is a multidisciplinary field cutting across health care, medical, and technological sciences, connecting medical informatics, business, and public health through internet-based technologies [7]. At the same time, experience has shown that the scalability and sustainability of mHealth services as part of an innovative digital health ecosystem could be hamstrung by factors such as stakeholder mismanagement, lack of political support, appropriate choice of technology, funding, and integration of mHealth into existing health programs in tandem with the Sustainable Development Goals [7].

In many developing countries, innovative initiatives aimed at enhancing the delivery of health services and disease management have been stifled by *pilotitis*—defined as a state of perpetual preliminary testing of projects in terms of which many technological innovations and initiatives, including mHealth, have not progressed to their intended full capacity [8]. A perennial state of *pilotitis* is caused by, among other factors, inadequate monitoring and evaluation systems, weak interorganizational and intraorganizational control mechanisms, and poor in-country digital architectures [8]. *Project Kopano* in South Africa and *Hello Mama* in Nigeria are examples of projects that stagnated because of pilotitis and did not develop beyond their pilot phases. Consequent to *pilotitis* and its negative organizational impacts, the objective of addressing Sustainable Development Goal 3 (good health and well-being) is jeopardized, as health technological developments are not accessible to most people for whom health care service provision is an absolute requirement [9].

The digital management of diseases through mHealth and eHealth reflects the development, adoption, and integration of ICT-based innovations in health care service planning, management, and delivery [10]. However, health care facilities and systems in many developing countries are still paper-reliant in many parts of their operations, which stifles progress insofar as digitally improving the quality, safety, and productivity of their health care services is concerned. mHealth has ushered in important changes through its facilitation of access and willingness to use portable devices for health care needs. mHealth is a medical and public health system that promotes the enhancement of health services with the support of wireless multimedia technologies such as mobile phones and other devices for monitoring patients and their recovery progress and well-being [11]. mHealth and eHealth have the potential to support and strengthen existing health care programs, rather than focusing on the discovery of new treatments and the development of clinical interventions by themselves [12]. Such a complementary approach in developing countries is significantly helpful, considering the plethora of factors such as adverse conditions (eg, poverty), limited investments, and resource constraints (eg, provision of safe drinking water, sanitation, basic education, medicines, and skilled personnel). For example, patients' adherence to medication could be enhanced through mHealth, rather than administering new

medication when the only issue was adherence (a factor of educating patients), and not a reflection on the effectiveness of the health care system in general or a particular program within the health system.

The application of mobile technology–based health care services (mHealth) is a credit associated with the introduction of ICT in health development. However, the design strategies and capabilities of such devices have also been questioned [11]. A differentiation between *user-centered* and *stakeholder-centered* designs is worth mentioning. In information technology systems architecture or design science, the 2 concepts are used interchangeably, although *stakeholder-centered design* is more applicable in the context of public health systems. In this regard, a stakeholder-centered design perspective was adopted in this study. A stakeholder design is defined as a process intended for the use of a (health) product at any moment and at any of its location points by all people during the product's lifetime to the advantage of those whose needs are not appropriately addressed [13]. Every effort should be made to identify and understand these stakeholders and address their needs to keep the chain intact [13].

The aim of this study was to share the perspectives of the different stakeholder groups involved in the design and implementation of a mobile app used in maternal health care services in a particular DHIE. The research question that guided the research for this paper is: What are the perspectives of the different stakeholder groups in the form of best practices, lessons learned, or issues identified based on their involvement in the design, development, and use of a specific mHealth application? The insights gained from this study should assist in moving health innovations beyond the pilot stage.

Background to the MomConnect Project of the NDOH

The MomConnect project is a product of the NDOH's National Digital Health Strategy (NDHS) intended to meet health targets and simultaneously maintain the momentum thereof, that is, scalability and sustainability [2,5]. The NDHS itself (launched in 2019) incorporated and combined eHealth and mHealth strategies, both of which have been periodically reviewed every 5 years since 2014 (the year of MomConnect's official launch). Conceptually, eHealth premises on the provision of health care through ICT to empower patients, families, and communities in the improvement, monitoring, and management of their health and well-being [11]. Meanwhile, mHealth relates to medical and public health practices and disease management programs supported by a range of interactive ICTs such as mobile phones, patient monitoring devices, and PDAs [1,14]. In essence, the current NDHS (2019-2024) is cognate from the review of the National eHealth Strategy (2012-2016), which focused mainly on improving governance systems and structures, integration of information systems, and technological enhancement and interface of the health care system and its users [6].

The merging of eHealth and mHealth by the NDOH signifies the need for critical considerations for a re-engineered and coordinated mHealth strategy that incorporates funding for eHealth and mHealth, impact assessment, management of data, effective leadership and governance from the NDOH, integration of lessons learned from other mHealth initiatives to avoid resource wastage and duplication of efforts, proactive evaluation of both mHealth and eHealth strategies, and change management and developing human resources for eHealth [5,6]. Moreover, contemporary issues of mHealth services could be addressed by applying digital development principles to strengthen best practices and address existing implementation gaps [10]. Furthermore, the centralization of a digital health system at the NDOH was viewed as a critical step in ensuring coordinated governance and effective leadership [6].

As espoused by the NDOH, the vision of the NDHS (2019-2024) is premised on the betterment of all South Africans' lives through digitized person-centered health services within an ecosystem inhabited by people and technology-driven processes [6]. The prioritized outcomes for better health for all South Africans rested on the key strategic pillars of person-centeredness, broadening access to health services, innovative sustainability, workforce-inspired economic development, and an interdepartmentally collaborative government approach. The strategic components of the digital health strategy are also mention worthy, as they cohere with some specific reference the study has made of the NDHS-MomConnect linkage, most of which have also emerged as critical variables in the findings. These 9 strategic components in the NDHS documents are leadership, stakeholder engagement, investment in strategy, governance, systems architecture and standards, digital apps and their services, connectivity infrastructure, legislation, policy and compliance, and workforce capacity (for economic development).

An Overview of MomConnect

The MomConnect project (which is not the strategy per se and is the demonstration case and overarching point of reference in this study) was officially launched in August 2014 as the NDOH's initiative for the improvement of maternal, child, and women's health (MCWH) services [6]. This project was founded on key elements, all of which are interstitially associated with maternal and child health and well-being until the newborn child is aged 1 year [5,6]. These elements are subscription (of users and beneficiaries to the mobile service), registration (of users on the national database through a common Unstructured Supplementary Service Data [USSD] number with the assistance of health care workers at the nearest health facility), SMS text messaging (NDOH SMS text messages relating to all relevant MCWH information), service rating (use of the free USSD number to rate the service using mobile devices), and compliments and complaints by users to local districts.

From its inception, this project was viewed as the *flagship* or standard of digitally propelled national programs of care for maternal and child health services in health care facilities across South Africa (Pillay Y, BP, unpublished data, April 2022) [5]. MomConnect was proudly welcomed by virtually all its beneficiaries, namely, the health care users who were mothers using MCWH services, pregnant women who came for antenatal care (ANC), and mothers who came for postnatal care. Notwithstanding some challenges, some observable success factors of MomConnect (since August 2014) included registration of >1.9 million users, a minimum daily rate of 1000 SMS questions asked by users, 9 times more compliments

received than complaints, and additional specific elimination of mother-to-child transmission messaging for women who have been HIV positive since September 2016 [6]. Nurses and auxiliary personnel at public health care facilities encourage and help pregnant women and new mothers register for MomConnect. Once registered, these women receive 2 to 3 SMS text messages per week based on their stage of pregnancy and through the child's first year of life. MomConnect message content is timed to the expected month of delivery and covers topics such as vaccination and checkup reminders, exclusive breastfeeding recommendations, psychosocial parenting tips, and baby development.

Given MomConnect's envisaged national impetus since its inception, it is therefore fait accompli that its scalability and sustainability factors would constitute the most salient variables within the DHIE parameters. The scalability and sustainability of a large-scale project such as MomConnect are achieved optimally by meeting targets and maintaining the momentum at the same time [2]. In general, project scalability and sustainability investigations are necessitated by challenges that were likely to relegate such potentially viable eHealth and mHealth services to a state of *pilotitis* [15]. Scalability most profoundly relates to quantitatively broadening or increasing access to health care services, whereas sustainability is concerned with qualitatively ensuring long-term success and continuation with unconstrained availability of financial, human, and infrastructural resources [5,8]. However, the current environment of MomConnect's implementation is still characterized by environmental factors reminiscent of a state of *pilotitis*. For instance, the implementation task team was focused more on scale (digital services expansion) than on sustainability (long-term durability) from the beginning of the initiative. It was performance driven, based on achieving targets. Once there was sufficient (quantitative) achievement on the target side, sustainability (qualitative) became a casualty. In addition, critical factors that were moderate at MomConnect's inception in 2014 became more critical, such as fundraising, exploring the registration of MomConnect as an independent entity, equitable allocation of contracts to competent digital service providers, stakeholder conflicts of interest when appointed to the board of the new entity, and perpetual legal advice being sought, among others.

The original MomConnect content was created by BabyCentre, adapted for the South African context by a team of local experts, customized for the length of an SMS text message (160 characters), and reviewed by a panel of experts including maternal health clinicians. A component of MomConnect has a task team that meets the most active stakeholders monthly [16]. MomConnect was selected as a case example in this study because it is the first nationally scaled-up mHealth service. An in-depth study of its implementation from a service design perspective may assist in obtaining evidence-based and best-practice mHealth implementation strategies. From a technical perspective, mHealth implementation also includes user-provider ethical components such as trustworthiness, privacy, and confidentiality.

As South Africa's first national-scale mHealth service, the implementation of MomConnect has the potential to fill the knowledge gap in mHealth implementation dynamics (which may include service design) [17]. There is currently limited published research in Sub-Saharan Africa on large-scale mHealth implementation on which to establish empirical investigation pertaining to the exploration of the well-designed efficacy and effectiveness of mHealth services [17]. The latter view is supported by the fact that the proliferation of mHealth initiatives in many developing countries has not necessarily translated into rollouts at the national level, as well as the practical implications of these projects on the routines of the facilities at which they have been rolled out [18] (NDOH, 2014).

Study or Research Framework

The study framework shown in Table 1 depicts 3 critical variables, namely, various categories of participants as the sources of data, research methods or instruments through which particular types or forms of data were generated, and the approach adopted to analyze the accumulated data. Necessarily so, these 3 diagrammatically represented variables are also indicative of the *route* of the study, that is, the most salient processes and activities performed throughout the study from its inception and conclusion [19,20]. The approach adopted by the researcher in analyzing the accumulated data in this predominantly qualitative-ethnographic study is critical and mention worthy, because it (approach) underpins the logic and rationale for the study and all its associated processes, which are captured to varying degrees of detail, particularly in the *Methods* and *Results* sections.

Regarding the aforementioned approach, the study framework reflects and encompasses a practitioner-researcher perspective in terms of which the researcher is simultaneously an active participant or observer in the situation being analyzed [21]. In addition, the practitioner-researcher approach has been influenced by the fact that the researcher is also a public health practitioner with >10 years of experience in the field. Moreover, the researcher served as a member of the MomConnect Task Team from November 2015 to June 2018 when he was employed by one of the implementing partner organizations and seconded to the NDOH to implement the elimination of mother-to-child transmission of the HIV component of the MomConnect initiative. This was the period during which the various critical activities and processes of the study were conducted.

Largely as a factor of the practitioner-researcher approach, the research approach depicted in Table 1 highlights the practitioner-researcher perspective as influential in shaping the study framework with respect to the chronology of the research process as clearly demarcated in the 2-phased stages (January 2017 to June 2018) showing the empirical and nonempirical (eg, document review) domains of the investigation. It is also clear from Table 1 that all the research-related activities and processes are cohesively bound and characterized by 2 indispensable components: simultaneity and contiguity.

Table 1. Depiction of study framework.

Stages of research and participant category and source of data	Activity period	Research method and type of data	Mode of data analysis and analytic tool
Phase 1			
MomConnect repository and archived data (including minutes and reports dated February 2014 to June 2018)	January 2017 to June 2018	Document and literature review	Content analysis
Phase 2			
Ministerial Advisory Committee on eHealth	January to December 2018	Interviews	Thematic analysis
MomConnect task team			
	January to December 2018	Stakeholder relationship mapping	Discourse and conversational analysis
	January to December 2018	Interviews	Thematic analysis
MomConnect task team meetings	January to June 2018	Ethnographic observations	Discourse and conversational analysis
Clinical staff	January to December 2018	Interviews	Thematic analysis
Auxiliary health personnel	January to December 2018	Interviews	Thematic analysis
Patients or health service users	January to June 2018	Focus group discussions	Thematic analysis

The notion of contiguity entails the inseparability of all relevant research variables, whereas simultaneity encompasses the development, occurrence, or undertaking of more than a single activity or process simultaneously [19,20]. For instance, all phase 2 processes and activities are contiguous with the core phase 1 activity. Inversely, the review of pertinent literature and relevant documents occurred throughout the research process. Thus, these 2 phases are contiguous on the basis of their complementarity as well. By contrast, all phase 2 activities and processes happened at different times and places in real time yet simultaneously in the context of the broader period during which they occurred and were performed.

Table 1 also clearly illustrates a multimodal approach of data collection and analysis, which justifies reference to convergent thinking and analysis as a tool for understanding the final outcomes and consequent findings, in terms of which the broader domain of MomConnect's implementation strategies could be justifiably assessed [20,22].

Methods

Research Setting

The study was conducted at 2 geographically disparate locations in the Gauteng Province, South Africa. The first setting was the most appropriate, given that it was the physical location for the offices of the NDOH, the fiduciary custodian of the MomConnect project. In addition, members of both the MomConnect Task Team and the Ministerial Advisory Committee on eHealth (MACeH) were more accessible at this site, as they regularly attended their meetings at the NDOH's offices. This venue was also crucial because it housed the archived documents in the MomConnect repository. These documents were instrumental in secondary data collection.

The second setting consisted of 4 inner-city primary health care (PHC) facilities (clinics) located in the largest city and economic hub of the country. These 4 clinics were selected because they were located in high-density populations, which ensured large-scale involvement of the selected participants. Furthermore, each of the 4 study clinics in Johannesburg's Region F offers similar PHC services that predominantly focus on HIV or AIDS.

Participant Characteristics and Their Recruitment

A total of 5 stakeholder categories were involved in this study, each representing both national-and facility-level perspectives. The first category, the MACeH and health facilities, was appointed by the Ministry of Health and is responsible for advising the Minister of Health on policy-related matters concerning eHealth. The second category, the MomConnect Task Team, involved representatives of different private and public sector organizations, academic institutions, and independent consultants. These representatives were involved at different stages of the MomConnect implementation and have held monthly meetings since the inception of the MomConnect project. The third category involved professional clinical staff at the 4 inner-city PHC facilities. Fourth, there were also auxiliary health personnel based at the same health care facilities. The fifth and final informant category consisted of patients or health care users. Table 2 illustrates the 5 stakeholder categories, as well as the primary data collection methods used for each.

Table 2. Participant categories sampled and their data collection methods (N=80).

Participants and stakeholders	Composition	Count, n (%)	Site and data collection method
Ministerial Advisory Committee on eHealth	Consists of 1 senior government official representative from each of the 9 provinces, academia, research organizations (eg, CSIR[a]), and private sectors, as per the government gazette	9 (11)	1-on-1 (face-to-face, telephonic, and virtual) interviews at NDOH[b] offices
MomConnect Task Team	NDOH officials and implementing partners (academia, funders, NGOs,[c] consultants, and research institutes)	15 (19)	1-on-1 (face-to-face, telephonic, and Skype) interviews at NDOH offices
Clinical staff	Professional nurses working at the 4inner-city clinics providing ANC[d] services in Johannesburg Region F	5 (6)	Face-to-face interviews at the inner-city PHCs[e]
Auxiliary health personnel	Staff based at the facility who are not registered clinicians but do interact with patients who come for ANC services, for example, health promoters, lay counselors, community health workers, and data capturers within the health care facility	6 (8)	Face-to-face interviews at the inner-city PHCs
Patients or users	Pregnant women and mothers visiting health care facilities for maternal, child, and women's health at the clinics; women who were at the facilities on the specific day when researchers were at the clinic were all sampled and formed part of FGDs[f]	45 (56)	Five focus group discussions of nine members each, at the inner-city PHCs

[a]CSIR: Council for Scientific and Industrial Research.

[b]NDOH: National Department of Health.

[c]NGO: nongovernmental organization.

[d]ANC: antenatal care.

[e]PHC: primary health care.

[f]FGD: focus group discussion.

Sampling of Research Sites and Participants

Purposive sampling was used for the selection of all key informants, based on our own professional judgment, experience, and knowledge of the research environment, informing us that the sampled participants complied with the requirements or criteria that we determined before the execution of the empirical data collection phase [20,23]. Nonprobability convenience sampling was used for the selection of the 4 inner-city research sites based on their ready availability and easy accessibility [9]. Both sampling methods were motivated by the fact that the researcher was a member of the MomConnect Task Team from October 2015 to June 2018. In this capacity, he was appointed as a digital health expert and seconded to the NDOH.

Description of the Different Research Methods

The MACeH representatives were selected mainly for policy-related reasons, given their knowledgeability and close interactions with the Ministry of Health. In contrast, the MomConnect Task Team members were sampled based on the fact that they represented different stakeholder constituencies involved in the implementation of the MomConnect project at the national, provincial, district, and subdistrict levels; attended the MomConnect Task Team meetings regularly; and were very knowledgeable about the project's functioning, mandate, and expected deliverables when appointed by the government. This team consisted of representatives of different private and public sector organizations, academic institutions, and independent consultants who have been involved at the different stages of the MomConnect project's implementation since its inception in 2014. Professional clinical staff were purposively selected based on their practice-related knowledge and work experience pertaining to the functioning of their inner-city health care facilities in Johannesburg. Meanwhile, the selection of auxiliary health personnel professional staff was influenced by our concern with perceptions of their exclusion in the MomConnect activities, although they were expected to render services related to the requirements of MomConnect users.

In contrast, the category of the purposively sampled health care service users were mothers who using MCWH services, pregnant women receiving ANC, and mothers visiting the facilities for postnatal care. These were the most direct beneficiaries of the government's MomConnect initiative, and their knowledge, experiences, and perceptions were indispensable to this study.

Data Collection Methods and Processes

Overview

In essence, this study was conducted in 2 phases that are distinguishable by their contiguity and simultaneity elements. In varying degrees, both Tables 1 and 2 provide significant information that also preludes the trajectory or approaches adopted for the data collection and analysis processes. It is worth noting that by virtue of the study's broader domain of simultaneity and contiguity, the data collection and its associated analysis processes have seamlessly (rather than chronologically) integrated a methods-based orientation for data collection and a participant-based orientation for the framework of results. Whereas the methods-based orientation inevitably emphasizes and accentuates the data, the participant-based orientation prioritizes and particularizes the source of the data itself. As such, relevant data were collected according to the research methods described below.

Literature and Document Review

The MomConnect repository itself was relevant, as it housed the official institutional memory of the MomConnect Task Team. This phase of data collection was of significant importance to the study, as it facilitated the evaluation of MomConnect's decision-making processes, providing a comprehensive background on the implementation process of the project, as well as an opportunity to examine the difference between the planned and actual implementation of the MomConnect project.

The document review included the official government's (NDOH's) MomConnect initiative work plan and other records. In this regard, the MomConnect repository, located in the NDOH, was also a reservoir of information, including the minutes of the MomConnect Task Team meetings. The MomConnect repository contains publicly available documents such as progress reports of implementation, archived and current minutes of task team project meetings, survey reports, and data and operational research documents or reports.

Stakeholder Relationship Mapping

Stakeholder relationship mapping basically refers to the identification and categorization of the main project participants (individuals, organizations, or institutions) who directly or indirectly have a vested interest in the ultimate outcome of the particular project based on their levels or stages of involvement in the very same project [24]. Stakeholder relationship mapping was of critical importance, especially because poor and weak program management challenges accounted for the failure of scalability and sustainability capacity required for the delivery of huge national projects such as MomConnect [25,26]. Therefore, stakeholder mapping was only applied to the participant category located within the decision-making and policy development echelons (such as the members of the MomConnect Task Team) rather than to the project implementers (eg, clinical and auxiliary health personnel) or end users (ie, patients at the PHC facilities). In this study, relationship mapping was only applied to the MomConnect Task Team members through a written exercise, filling-in an informed consent form, and returning to the researcher on the same day. After informed consent was obtained, task team members were emailed an exercise used to determine the nature and range of their past and current relationships that may have some impact on the MomConnect project.

Having obtained considerable background information and knowledge through the systematic review of relevant literature and documents, stakeholder mapping (similar to research participant selection criteria) constituted the logical phase before the actual empirical data collection itself through interviews and focus group discussions [27]. This mapping of the various groups and stakeholders was critical as it enhances more understanding of the relationships and interrelatedness of individuals, groups, and the specific mHealth service itself. The mapping process is focused on the exploration and understanding stakeholder relationships rather than on finding out *who* they are [28].

Semistructured Face-to-face Interviews

This phase of data collection was conducted with the MomConnect Task Team members, MACeH, and clinical and auxiliary health personnel.

Ethnographic Observation of MomConnect Task Team Meetings

In research involving an empirical component, the observation of participants is an ongoing process (Pillay Y, BP, unpublished data, April 2022). From an ethnographic perspective, the observation of research participants further provides an opportunity for the researcher to interact directly in conversations or dialogues with the participants and observe their attitudes, behavior, and interaction toward each other and one another within their ecological parameters of the natural environment to which they are very familiar and in which they interpreted their reality and conditions [29,30]. In this regard, participant observation complemented both the primary data collection methods (ie, individual interviews and focus group discussions).

Participant observation was facilitated by means of the researcher's physical observation of and listening to MomConnect Task Team members in their monthly meetings. The observation rationale was to add value to this study to the extent that more understanding was important for assessing the stakeholder relationships and collaboration. The latter 2 aspects were critical factors because they provided a basis for the relative determination of the success or failure of project design, planning, and implementation processes and dynamics at both the bureaucratic and technocratic levels, given the composition of both the MACeH and the MomConnect Task Team [19,31]. For the purpose of this study, and given the heterogeneous representation of interests and constituencies within the MomConnect Task Team, 3 of their monthly meetings were attended. Field notes were taken, focusing on their interpersonal relationships, as well as their decision-making processes and procedures in meetings. This phase was complementary to the review of relevant MomConnect repository documents.

Focus Group Discussion With Service Users at Inner-City PHC Facilities

As shown in Table 1, this targeted form of engagement, interaction, and conversations was conducted with pregnant women who came for MCWH services and ANC and mothers who came for postnatal care. Focus group discussions were advantageous in that virtually all participants were more at ease sharing and learning in a group from the experiences, knowledge, and perceptions of others [32].

Data Analysis Approaches

Both the heterogeneous nature of participant categories and the range of ethnographically oriented instruments of data collection triangulated the data analysis approach involving content, thematic, discourse, and conversational analysis. The study emphasizes that, despite their terminological variation, the modes of analysis all fundamentally focus on deriving meaning from themes developed from the content of empirical and nonempirical research methods used in this study [18].

Ethics Approval

Ethics approval was obtained from Cape Peninsula University of Technology. Furthermore, access to conduct research was endorsed by the NDOH, Gauteng Department of Health and City of Johannesburg.

Results

Overview

In contradistinction with the methods-based approach of data collection, the results framework presented in Table 3 is emblematic of a participant-based orientation. As such, the results or findings are presented based on the type or category of participants as providers or sources of the self-same data obtained through different methods. Notwithstanding the transcendence of participant categories over the different data collection methods in this regard, both methods and participants are still contiguously linked. There could be no empirically generated data without participants as the vital source of the information sought to fulfill the study's aim [1,20]. Accordingly, the results framework in Table 3 is the outcome of the participant-method contiguity. Most importantly, the results reflect the convergence of data analysis methods, that is, the predominant thematic mode complemented by conversational and discourse analysis.

Owing to the variability of research methods applied and the vastness of the data collected, the results are presented in varying degrees of detail consonant with aspects they address in relation to the development, implementation, and usability of the MomConnect system of digitizing health care services *for all South Africans* [6]. From the perspective of this study, the development-implementation-usability continuum is pivotal in determining the extent to which the NDOH's MomConnect initiative could be demonstrated as a case of policy or strategy success or failure with respect to its attendant sustainability and scalability factors.

Table 3 is reflective of the eventual outcomes of various data analysis processes adopted (Table 1) to construct meaning from various participant categories representing various components and aspects in the MomConnect policy development, strategy design, and implementation and health care service users' benefit. In the end, the convergence of the various data analytic modes also reflects the inextricability of participants' environmental dynamics (eg, vested interests and influences) and the inevitable researcher-practitioner approach adopted. Owing to the vastness of the data collected, the emergent themes have been paired globally in groups rather than individually.

Table 3. Results framework.

Participants and main theme	Main category	Subcategory
Ministerial Advisory Committee on eHealth (interviews)		
Governance and leadership	mHealth[a] centralization	mHealth service rationalization
Strategy integration	eHealth and mHealth strategy perceptions	Strategy application feedback
Stakeholder involvement	Clinicians, technophobia or capacity building; mHealth providers and consumer engagement	—[b]
Research and development	Demonstrable evidence of implementation outcomes or impact; mHealth piloting perspectives	—
Service continuity		
Sustainability		Total cost of ownership and co-utility; cost of ownership and cost utility; outsourcing culture
Scalability		NDOH[c] financial and human resources; provincial realities regarded as mHealth barriers; computing infrastructural issues
Design thinking		Lessons learned
Service implementation		Human and financial resources
Ecosystem or environment		
Organizational		Vision, policies, and guidelines; governance and leadership; political authority or oversight
Ethical aspects		Privacy and security; data ownership
Integration		Technical: infrastructure and interoperability
MomConnect Task Team (minutes)		
Service conceptualization		
Stakeholder considerations		Facility-level consultation and collaboration
Design process		Considerations; research and expansion considerations
Integration		Technical: infrastructure and interoperability
Roll out	National to provincial scaling-up process; operations and performance	—
Service continuity	Service continuity; sustainability and evolution	—
MomConnect task team (interviews)		
MomConnect as a case example		
	Member relationship mapping	—
	Support for MomConnect as a case example; differing piloting views	—
Critical considerations	mHealth and eHealth strategy	Life span within NDOH or integration of initiative within health programing; ethical service implementation; uncertainty over sustainability
Ministry of Health prerogatives	Leadership and management; teamwork; operations; recommendations	User-centered design; sustainability; change management; stakeholder management
Facility-level (clinicians, auxiliaries, and service users)		
Service touch point capacity		
	Stakeholders	Involvement of nurses or clinicians (capacity building and NurseConnect); mothers, pregnant women, and caregivers; foreign nationals (challenges)

Participants and main theme	Main category	Subcategory
	Service implementation	Content of information; ethical considerations; MomConnect helpdesk and interactive communication with nurses; subscription and marketing; service rating or feedback
	Health care facility environment	—
	Operations	—

[a]mHealth: mobile health.

[b]Not available.

[c]NDOH: National Department of Health.

Findings Emanating From the MACeH

A total of 6 main thematic statements and their associated categories and subcategories emerged and are discussed in the next sections.

Governance and Leadership

Especially in developing countries, governance and leadership issues have been major determinants of the sustainable implementation of health programs and services [25]. In this regard, the MACeH interviews reflected general consensus for the support of centralizing mHealth governance and control in the NDOH: "In itself, the centralization factor entails both the pros and cons of leadership and governance issues." In this case, centralization was viewed as beneficial because there were a number of unsupervised health initiatives throughout the country, and strong recommendations were mooted for a national database or register to track the performance of mHealth in various parts of the country. Such a step was viewed as advancing the issues of both scalability and sustainability from the very inception of any health care initiative by the Department. This was also viewed as a transparent process that would reduce wastage and duplication of resources. Such a trajectory would ensure that different mHealth implementers are fully acquainted with what other stakeholders are doing. Consequently, mHealth implementers would be capacitated to build on what already exists and learn from the experiences of others. Notwithstanding its advantages, some disadvantages of centralization were also observed. The bureaucratic nature of a centralized system implied that implementers take a long time to obtain permissions, which causes an incremental loss of opportunities during delays [12]. In addition, little or no centralization inadvertently encouraged everyone everywhere to have their own few mHealth projects that never scaled up and remained mostly unknown to the NDOH. In such instances, the need for centralization could be balanced with the need for the devolution of authority, for example, by decentralizing certain implementation functions and regulatory authority to the provinces, regions, and districts [23]. Rationalization relates to the justification of a course of action associated with any aspect of the MomConnect system.

Strategy Integration

MomConnect is a product of the overall digital health strategy of the NDOH, and the finding is that the mHealth strategy is not sufficiently integrated into either the eHealth strategy or the health strategies of the country: "I think through that we are missing an opportunity to ensure that the various types of mHealth that is being practiced in the country is fully aligned with the health transformation work that is being led by the ministry and department." An mHealth strategy needs to be supported by a solid implementation strategy.

Stakeholder Involvement

Clinicians were not effectively involved in the system's design: "...the health system is completely separate from the PHC system." For instance, the scant involvement of nurses could stereotypically portray them as *technophobic* and stifle their capacitation and competence [3,8].

Research and Development

Research on digital health is crucial, considering that mHealth is growing constantly and continuously as part of eHealth and is a component of all aspects of health (prevention, diagnosis, treatment, and research) [7,9]: "...at the time MomConnect was rolled out there was not enough substantiated evidence to support national roll out." It was regarded as a politically driven project: "...we didn't talk at all around research which I think is an ongoing real problem."

Service Continuity

Service continuity is a direct reference to the long-term duration of any program or project [10,13]. In considering the scalability and sustainability factors, the MACeH was also guided by weighing costs associated with sole or co-ownership and outsourcing of certain aspects such as whether funding has been planned for continuing with mHealth projects past the DHIE stage. There needs to be a process of assessing the sustainability of the mHealth initiatives by the relevant stakeholders to ensure that all risks have been evaluated. There is a need for standards, interoperability, and human and financial resources to render the service implementable. Women will be empowered through the MomConnect project, resulting in them having information expectations that could then place a burden on clinics and hospitals to also meet that expectation:

> *The NDoH uses consultants, that consultant has a life span at the department, you can only employ that person for so long. If you want to be scalable and sustainable you must write a standard approach that is not vendor biased and can be supported by any software vendor.*
>
> *How can we sustain anything if 50% of it is from donors.*

Ecosystem or Environment

A digital ecosystem encompasses the entire community of networks within the digital environment to establish best practices [1]. The NDOH was the central authority to safeguard MomConnect users from, among others, protecting their privacy: "Start looking at security issues that may influence the project. [Informed] Consent and POPI [Protection of Personal Information Act of 2013]..." There may be concerns regarding confidence in the system [3]. Scalability and sustainability should guide implementation at the provincial and district levels.

Findings Emanating From MomConnect Task Team Minutes

The themes are discussed in the next sections.

Service Conceptualization

Service conceptualization involves developing and designing a program (at the theoretical or planning stages)—in this case, the MomConnect initiative—such that its technological or digital relevance to and application by the intended users or beneficiaries achieves the required benefits of broadening health care services to all South Africans (Pillay Y, BP, unpublished data, April 2022) [33]. The MomConnect service conceptualization was captured in this study in terms of the affected stakeholders, facility-level consultation and collaboration, and the acceptability (or otherwise) of the design process value chain. During the conceptualization of the service, stakeholders were not only confined to nursing staff: From the afore-cited minutes, it is evident that effective stakeholder considerations extended beyond the nursing personnel as implementers at the facilities. Such a conceptual orientation stood the MomConnect service's implementation in good stead, as collaboration and cooperation were critical for health care service delivery [34].

Design Process

The process of developing and designing a health service requires an in-depth understanding and knowledge of the field, as well as consideration of multidisciplinary perspectives [8,13]:

> *...research with nurses to improve user-centred design...They [mobile service designers and providers] are also going to do research that will allow us [Task Team/policy makers] to understand which handsets our clients are using...back-end system, mobile operators, service providers, messaging subgroup rep, registration process, communication strategy, launch, field testing...If we design the system properly, we might be able to have them [patients] enter data in the waiting room.*

The fact that MomConnect is still in existence in all 9 provinces despite initial imbalances bears testimony to extrinsic factors such as the role played by reputable service designers and providers outside of the NDOH policy-making value chain [32,34].

Integration

The willingness to integrate the MomConnect initiative and not implement it as a silo was recorded. However, there were no documented indications of any other health service initiatives to which MomConnect could be integrated. MomConnect was initially conceived as a standalone digitized health service. Only data integration was explicitly mentioned in terms of linking MomConnect data to existing health information system (HIS) data for reporting. The South African ID has been integrated into biometric ID systems that are compliant with various forms of DNA analysis and has been proposed as an identifier for the system. However, foreign national ID numbers must also be captured. The system was also unable to capture dates of birth for underaged pregnant mothers who did not have an ID or foreign nationals who did not have their passports when registering:

> *To allow for maximum interoperability there should be a number of ID types which will allow all people eligible to receive care to be registered. In South Africa, everyone has a right to emergency medical treatment. Therefore, identifiers which cover foreign nationals, refugees and people without documents should be used.*

Infrastructure and Interoperability

The technical aspects of mHealth are the quintessential reflection of the integration or interface of ICT systems for human services [15]. The technological infrastructure required for connectivity at a national scale was enhanced by strong NDOH partnerships with established mobile phone operators to boost the economies of scale and size, suggesting that it is technically possible to deliver mHealth interventions to large populations at low cost because downloading and automation to send SMS text messages can be once-off processes [26,35,36].

Rollout

The proliferation of mHealth initiatives in many developing countries has not necessarily translated into rollouts at the national level, considering the practical implications of these projects on the daily routines of the facilities at which they have been rolled out [18]. Therefore, the capacity of the NDOH to roll out MomConnect as a national service to its intended beneficiaries was an important determinant of whether this service could be implemented on a large scale [25]. In the context of the MomConnect Task Team minutes, rollout and expansion of the service were both geographical (throughout all 9 provinces) and technological (operations and performance). There were efforts to prevent the MomConnect project from another project that remains in the pilot state:

...[the Health] minister has announced [on 30 July 2014] that he will be embarking on road shows to introduce MomConnect to the health professional in each province. Demonstration sites were selected, communication materials were approved and translated into other official languages. Establish an ongoing system of formal evaluations and review of potential mHealth projects.

Service Continuity

Determining service continuity was necessary to establish measures recorded in the minutes to ensure the survival of MomConnect beyond its official launch in August 2014. Other

than their involvement in providing their network infrastructure for MomConnect use, the role of established mobile telephone companies was notably recognizable, facilitating the USSD service to ensure that the service is not only scalable but also technologically viable to continue to the present [16,33]:

> *It was suggested [in the meeting] that one should look for risk sharing agreements that create long term possibilities and business models. Look into ways to make MomConnect more sustainable and generate funds.*

Findings Emanating From MomConnect Task Team Interviews

Whereas the previous section focused on analyzing data from indirect engagement with the participants (MomConnect Task Team members), the current section premises on directly obtained data with the self-same participant category through the individual interviews. These themes are discussed in the next sections.

Support for MomConnect as a Case Example

Support for an innovative digital health strategy by relevant stakeholders is critical for its long-term sustainability [2,5]. In the case of MomConnect, there was significant support at the different levels of the implementation value chain:

> *It's important, if eHealth has to be implemented there has to be somebody with clinical background, who understand the technicalities required to support...Everybody realized this was NDOH project and were keen to be involved.*

Differing Piloting Views

It is expected that MomConnect content should be in line with, and not deviate from, maternal health guidelines as a contribution to advancing the goals of the country's maternal health program. However, participants had different views on the use and application of MomConnect as an incipient digital health project of the NDOH through its digital HIS strategy. Concerns were raised about the likelihood of a disjuncture between strategy or policy development and implementation:

> *I have worked with both eHealth and mHealth strategies outside MomConnect in my previous role...My general assessment about those documents [guidelines] is that they are very theoretical and difficult to implement in real terms, especially the eHealth strategy...[it] is very high-level and doesn't really explain concretely how to do the things that are recommended.*

Although it emphasizes the perspective of incongruences, participants raised concerns about the political overtones of the environment within which the DHIE exists. Such a politicized context could be detrimental to the sustainability of the MomConnect project when politicians subsume the roles of experts, implementers, and other practitioners [14].

Possible reasons for the different responses were that some participants were involved in the drafting of the strategy because they have been working with the Department on different eHealth initiatives and knew what the strategy entails. There were also participants who felt that their roles were very operational and did not need any reference to the strategy. There was also a response that an international organization visited the MomConnect office to benchmark the initiative, and it was difficult to explain the relationship between the initiative and the strategy because the particular organization expected a clear link between the strategy and the initiative. The participants further intimated that there was no indication that MomConnect was adequately responsive to a viable eHealth or mHealth strategy. It was perceived more as a high-level initiative of the Department and replete with policy *speak* with no significant evidence of a sustained national rollout other than addressing the political egos of those connected to the highest echelons of power in the Department:

> *I have worked with both eHealth and mHealth strategies outside MomConnect in my previous role...My general assessment about those documents [guidelines] is that they are very theoretical and difficult to implement in real terms, especially the eHealth strategy...[it] is very high-level and doesn't really explain concretely how to do the things that are recommended.*

Critical Considerations

It emerged during the interviews that the MomConnect Task Team critically considered a range of issues, most prominently the life span of MomConnect within NDOH (its integration within health programming), ethical service implementation, and uncertainty over its sustainability. There were views that this initiative was not integrated with other programs within the Department and that it was implemented as a solo high-profile initiative of the former health minister. The issues around integration included, but were not limited to, data, compliance with HIS reporting, and data management. For instance, the HIS unit has specific rules, but they were perceived to be not fully applied by the MomConnect team:

> *It has to be integrated within the provincial maternal services. MomConnect is seen as separate entry and is not part of, look at the two units, HIS and HIV/AIDS.*
>
> *...one of the key issues with any strategy in the space of health technology...is transforming...we have a health system that is changing...having a strategy of five years is actually too little...*

Ministry of Health Prerogatives

The Ministry of Health is the de jure custodian of the MomConnect digital system, and as such, certain fiduciary responsibilities and expectations are executed through the NDOH [6]. Chiefly, these include leadership, change and stakeholder management, operations, and user-centered design. Collectively, these prerogatives are indicative of attempts to ensure sustained longevity of the overall health digital innovation strategy of the Department [5].

Leadership and Change and Stakeholder Management

Some concerns were raised regarding the management of stakeholders:

> *To have leadership from the NDoH is key. MomConnect has been good at leadership coming from the department. Nobody does anything unless the NDoH signed on it.*

On change management, the following excerpt captures the unanimous views of the interviewees in this regard:

> *There are no contracts. All partners are taking some level of risk, no formal agreement to say "this is how we will sustain the programme going forward. We need proper project management and change management strategy."*

As opposed to the policy-level sites of the MACeH and MomConnect Task Team, the health care facilities (especially) at the local and district levels are at the coalface of the MomConnect system's implementation [6,33].

Clinical Staff (Nurses)

Staff expressed that they felt excluded and alienated, as they were not part of MomConnect development. They were only told to register and subscribe patients to it, which was a nonclinical (auxiliary health) function. Nurses reported that the content of the SMS text messages received by the women was never shared with them. They only heard from patients when they were asked for clarity about specific messages or when they confirmed specific elements of ANC. Some nurses even reported subscribing to MomConnect for the purpose of familiarizing themselves with the contents of MomConnect on the end users' mobile phones. Because nurses do not usually register patients on MomConnect, the counselors and health promoters complained about the additional workload caused by the time taken to register users on the MomConnect platform:

> *I only heard of MomConnect 4 weeks ago when I started in antenatal care. I do not know the content of the messages our patients receive. We need to get the messages so that the content can be used as part of health education.*
>
> *...our biggest challenge is language barrier as most of our clients are from out of our South African borders...due to language barriers, patients don't receive information...*

Mothers and Pregnant Women

Disgruntlement was expressed by the pregnant women regarding the technicality and language barriers posed by the MomConnect functionality on their mobile devices. The disgruntlement emanated largely from foreign nationals, whose home languages were not part of South Africa's 11 official languages. As such, they (foreign nationals) frequently experienced communication barriers and understanding of MomConnect content. All pregnant women (regardless of country of origin) also experienced technical problems due to connectivity and time-out sessions induced by USSD operational failures in the provider network system. Notwithstanding, virtually all participating pregnant women expressed their satisfaction with the MomConnect SMS text messages and commended the service on the content regarding nutrition and preparing for the labor trip, and the nontechnical language used was easy to understand. However, there was a preference for the use of English as opposed to their own vernaculars because the translation in their vernacular was perceived as bombastic.

Service Implementation

The implementation and national rollout of a digital health service the size of MomConnect is largely dependent on the broad involvement and capacity of stakeholders at the health care facilities as the points of service [9,10]. Accordingly, the (clinical and nonclinical) operational context or environment of the health care facility and its units also determine the level at which the service is offered, whether it is acceptable or satisfactory or otherwise [2,8]. The findings indicated a general level of acceptable service implementation.

Discussion

Principal Findings

The principal findings in this study collectively reflect the interconnectedness between the adopted research approach (Table 1), the varied participant categories (Table 2), and the framework within which these findings were formulated (Table 3). The practitioner-researcher approach facilitated the seamlessness of the thematic, discourse or conversational, content, and convergence analytical modes. By virtue of his knowledge and experience as a former MomConnect Task Team member and based on the period during which the study was conducted, the findings are then the product of "...research in practice, a close study of practices with access to pertinent data" [21].

The study's most fundamental purpose was to use a stakeholder perspective as a preferred method to examine the scalability and sustainability factors of the NDOH's current MomConnect initiative, from which policy guidelines could be developed to inform on possible areas or aspects requiring improvements. A close observation of the results framework clearly indicates that the findings are located within 2 principal aspects of an innovative digital health ecosystem, namely, the digital environmental factors and the practice-related implementation attributes.

Regarding legislative and policy compliance, there are laws in South Africa, but the implementation of mHealth (and other spheres of public life) has not been consistently enforced. This study did not specifically examine law and health information. However, from the data generated in this study, there is an applicable regulation to mHealth, and implementers are aware of its applicability on mHealth. From the themes, particularly from the facility-level implementers as the service point workforce, a curriculum for eHealth was proposed. However, there is no reported process for the proposed curriculum to be developed or under review.

From a clinical perspective, the effectiveness of mHealth interoperability relies on a partnership between patients and clinicians' workflows. In particular, mHealth devices would benefit from interoperability standards to ease integration with

other health software apps, which are increasingly required to organize the large amounts of data collected [12,15]. In the context of this study, the digital HISs served as the context for MomConnect standards and interoperability. Information regarding full implementation following industry standards and usability across various devices and platforms showed little coverage and application.

Network infrastructure development and readiness were largely facilitated through partnerships with established mobile technology operators, as the NDOH does not have its own. This is an instance of the worth and usefulness of private-public partnerships, especially in a sphere of public life, where technological developments necessitate capital cost sharing (Pillay Y, BP, unpublished data, April 2022) [12]. For network services and device applications, the study confined itself more to the end-user features than to the system design aspects. South Africa's list of network facilities includes, but is not limited to, geographic information system mapping.

The objective of this study was to review the NDOH's MomConnect initiative as a case study of an existing government mHealth strategy, followed by the design of an improved strategy based on best demonstrated practices (considerations and methods) and learned experiences from the perspectives of multiple stakeholders within the DHIE in South Africa.

Strengths and Limitations

This study is strengthened by its user- and stakeholder-focused orientation, which accommodated both the nonprobability judgment and convenience sampling strategies. Many studies in this field are more technology or device-oriented and focus on *high-end* users of mobile technology. In contrast, the study could have a limited reach insofar as its confinement to PHC users in a metropolitan inner-city area, which might have different comparable outcomes in rural, suburban, and condensed informal settlement contexts. However, a significant aspect of the broader limitations was mitigated by the heterogeneity of the sampled participant categories representing different *vested* interests in society.

Future Research Directions

Further research is needed for the retrospective analysis of mHealth initiatives and the forecasting of the sustainability of current and future mHealth initiatives in South Africa, that is, the cumulative effect and impact of mHealth strategies and projects with similar or more emphasis accorded by the NDOH. In addition, more research is required from public health practitioners in practice on mHealth, as they were in daily contact with the mHealth beneficiaries (end users), which aptly placed them to obtain first-hand stakeholder perspectives and experiences [9,20].

Conclusions

The fact that MomConnect is generally a usable technology-based communication system does not preclude the identification of alternative interventions to ameliorate or even radically improve its shortcomings. For instance, the current caveat emptor approach renders pregnant women vulnerable to the same challenges generally experienced by most South Africans with internet-based broadband distribution. Notwithstanding the information distribution efficiency for antenatal and postnatal purposes, the USSD mechanism still allocates a cost to the user, albeit less than the sometimes prohibitive costs of data compared with the rest of the world. Therefore, it is unsurprising that the main critical considerations were funding for eHealth, including mHealth. Sustainability is a concept that must be considered throughout the implementation process of the MomConnect project. However, there was less consideration at the beginning of this initiative because we could consider sustaining something that has not even scaled. Funding for eHealth, including mHealth, is critical. The government must have its own funding mechanisms and must not depend only on funders.

Acknowledgments

Academic supervisors RDLH and PN are acknowledged for their academic supervision and guidance. The National Department of Health is also acknowledged for endorsing the study of INS in the MomConnect initiative.

Conflicts of Interest

None declared.

References

1. Iyawa GE, Herselman M, Botha A. Digital health innovation ecosystems: from systematic literature review to conceptual framework. Procedia Comput Sci 2016;100:244-252. [doi: 10.1016/j.procs.2016.09.149]
2. Schneider B, McDonalds S. Scale-Up in Education: Ideas in Principle (Volume 1). Lanham, Maryland, United States: Rowman & Littlefield Publishers; 2006.
3. Jeffries M, Phipps D, Howard RL, Avery A, Rodgers S, Ashcroft D. Understanding the implementation and adoption of an information technology intervention to support medicine optimisation in primary care: qualitative study using strong structuration theory. BMJ Open 2017 May 10;7(5):e014810 [FREE Full text] [doi: 10.1136/bmjopen-2016-014810] [Medline: 28495815]
4. Silva BM, Rodrigues J, de la Torre Díez I, López-Coronado M, Saleem K. Mobile-health: a review of current state in 2015. J Biomed Inform 2015 Aug;56:265-272 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.003] [Medline: 26071682]

5. Pillay Y, Motsoaledi PA. Digital health in South Africa: innovating to improve health. BMJ Glob Health 2018 Apr 24;3(Suppl 2):e000722 [FREE Full text] [doi: 10.1136/bmjgh-2018-000722] [Medline: 29713513]
6. National Digital Health Strategy for South Africa 2019 – 2024. Pretoria: National Department of Health Republic of South Africa; 2018.
7. Ali EE, Chew L, Yap KY. Evolution and current status of mhealth research: a systematic review. BMJ Innov 2016 Jan 05;2(1):33-40. [doi: 10.1136/bmjinnov-2015-000096]
8. Fanta GB, Pretorius L. A conceptual framework for sustainable eHealth implementation in resource-constrained settings. South African J Industrial Eng 2018 Nov;29(3):132-147. [doi: 10.7166/29-3-2055]
9. The need for implementation science research in mhealth: a vehicle from pilot to scale-up. The University of British Columbia. URL: http://id.med.ubc.ca/2017/07/01/the-need-for-implementation-science-research-in-mhealth-a-vehicle-from-pilot-to-scale-up/ [accessed 2022-05-02]
10. Pankomera R, van Greunen D. A model for implementing sustainable mHealth applications in a resource-constrained setting: a case of Malawi. E J Info Sys Dev Countries 2018 Feb 06;84(2):e12019. [doi: 10.1002/isd2.12019]
11. Whittaker R. Issues in mHealth: findings from key informant interviews. J Med Internet Res 2012 Oct 02;14(5):e129 [FREE Full text] [doi: 10.2196/jmir.1989] [Medline: 23032424]
12. Fleming JN, Taber DJ, McElligott J, McGillicuddy JW, Treiber F. Mobile health in solid organ transplant: the time is now. Am J Transplant 2017 Sep 17;17(9):2263-2276 [FREE Full text] [doi: 10.1111/ajt.14225] [Medline: 28188681]
13. User centered design is dead. m3 design. URL: https://www.m3design.com/user-centered-design-is-dead/ [accessed 2022-05-02]
14. Bull S, Ezeanochie N. From foucault to freire through Facebook: toward an integrated theory of mHealth. Health Educ Behav 2016 Aug 17;43(4):399-411. [doi: 10.1177/1090198115605310] [Medline: 26384499]
15. Fotso JC, Crawford J, Noordam A. Strengthening health systems for improved access to maternal, newborn and child health services: why mHealth should be part of the solution, and why rigorous, theory-based evaluations matter. In: Proceedings of the 7th African Population Conference. 2015 Presented at: 7th African Population Conference; Nov 30-Dec 4, 2015; Johannesburg, South Africa URL: https://uaps2015.princeton.edu/abstracts/150098
16. Sawubona MAMA : using mHealth to improve maternal, neonatal and child health outcomes in South Africa. Karolinska Institutet. URL: https://openarchive.ki.se/xmlui/handle/10616/46402 [accessed 2022-05-02]
17. Coleman J, Bohlin KC, Thorson A, Black V, Mechael P, Mangxaba J, et al. Effectiveness of an SMS-based maternal mHealth intervention to improve clinical outcomes of HIV-positive pregnant women. AIDS Care 2017 Jul 20;29(7):890-897 [FREE Full text] [doi: 10.1080/09540121.2017.1280126] [Medline: 28107805]
18. Chib A, Wilkin H, Hoefman B. Vulnerabilities in mHealth implementation: a Ugandan HIV/AIDS SMS campaign. Glob Health Promot 2013 Mar 02;20(1 Suppl):26-32. [doi: 10.1177/1757975912462419] [Medline: 23549699]
19. Leon N, Schneider H, Daviaud E. Applying a framework for assessing the health system challenges to scaling up mHealth in South Africa. BMC Med Inform Decis Mak 2012 Nov 05;12(1):123 [FREE Full text] [doi: 10.1186/1472-6947-12-123] [Medline: 23126370]
20. Creswell JW, Creswell JD. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, 5th edition. Thousand Oaks, California: SAGE Publications; 2017.
21. Goldkuhl G. The research practice of practice research: theorizing and situational inquiry. Int J Commun Inf Technol Work 2011;5 [FREE Full text]
22. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: implications for conducting a qualitative descriptive study. Nurs Health Sci 2013 Sep 11;15(3):398-405. [doi: 10.1111/nhs.12048] [Medline: 23480423]
23. Brink H, van der Walt C, van Rensburg G. Fundamentals of Research Methodology for Healthcare Professionals, 3rd edition. Cape Town: Juta Legal and Academic Publishers; 2012.
24. Fiordelli M, Diviani N, Schulz PJ. Mapping mHealth research: a decade of evolution. J Med Internet Res 2013 May 21;15(5):e95 [FREE Full text] [doi: 10.2196/jmir.2430] [Medline: 23697600]
25. Mawela T, Ochara N, Twinomurinzi H. E-government implementation: a reflection on South African municipalities. South African Comput J 2017 Jul 08;29(1) [FREE Full text] [doi: 10.18489/sacj.v29i1.444]
26. Miah SJ. Large-scale innovations and approaches for community healthcare support in developing nations. In: Khan SU, Abbas A, Zomaya AY, editors. Handbook of Large-Scale Distributed Computing in Smart Healthcare. Cham: Springer; 2017.
27. Serrano A, Garcia-Guzman J, Xydopoulos G, Tarhini A. Analysis of barriers to the deployment of health information systems: a stakeholder perspective. Inf Syst Front 2018 Jun 30;22(2):455-474 [FREE Full text] [doi: 10.1007/s10796-018-9869-0]
28. Science, technology, innovation and partnerships for development: an evidence gap map. International Initiative for Impact Evaluation. URL: https://3ieimpact.org/evidence-hub/publications/evidence-gap-maps/state-society-relations-low-and-middle-income-countries [accessed 2022-05-02]
29. Kumar R. Research Methodology: A Step-by-Step Guide for Beginners, 3rd edition. Thousand Oaks, California, United States: SAGE Publications; 2014.
30. Yin RK. Qualitative Research from Start to Finish, 2nd edition. New York: The Guilford Press; 2015.

31. Kuziemsky CE. Enhancing 'Fit' of Health Information Systems Design Through Practice Support. Pennsylvania, USA: IGI Global; 2008.
32. Lin M, Hughes B, Katica M, Dining-Zuber C, Plsek P. Service design and change of systems: human-centered approaches to implementing and spreading service design. Int J Design 2011;5(2):73-86 [FREE Full text]
33. Barron P, Peter J, LeFevre AE, Sebidi J, Bekker M, Allen R, et al. Mobile health messaging service and helpdesk for South African mothers (MomConnect): history, successes and challenges. BMJ Glob Health 2018;3(Suppl 2):e000559 [FREE Full text] [doi: 10.1136/bmjgh-2017-000559] [Medline: 29713503]
34. Carlsson SA, Henningsson S, Hrastinski S, Keller C. Socio-technical IS design science research: developing design theory for IS integration management. Inf Syst E Bus Manage 2010 Oct 14;9(1):109-131. [doi: 10.1007/s10257-010-0140-6]
35. Bateman C. Using basic technology--and corporate social responsibility--to save lives. S Afr Med J 2014 Dec 13;104(12):839-840. [doi: 10.7196/samj.8992] [Medline: 26042259]
36. Shukla SN, Sharma J. Potential of mHealth to transform healthcare in India. J Health Manag 2016 Sep 16;18(3):447-459. [doi: 10.1177/0972063416651589]

Abbreviations

ANC: antenatal care
DHIE: digital health innovation ecosystem
HIS: health information system
ICT: information and communication technology
MACeH: Ministerial Advisory Committee on eHealth
MCWH: maternal, child, and women's health
mHealth: mobile health
NDHS: National Digital Health Strategy
NDOH: National Department of Health
PHC: primary health care
USSD: Unstructured Supplementary Service Data

Edited by L Buis; submitted 10.02.20; peer-reviewed by V Horner, M Herselman; comments to author 12.03.20; revised version received 02.05.20; accepted 20.11.20; published 16.06.22.

Please cite as:
Sibuyi IN, de la Harpe R, Nyasulu P
A Stakeholder-Centered mHealth Implementation Inquiry Within the Digital Health Innovation Ecosystem in South Africa: MomConnect as a Demonstration Case
JMIR Mhealth Uhealth 2022;10(6):e18188
URL: https://mhealth.jmir.org/2022/6/e18188
doi:10.2196/18188
PMID:35708756

Original Paper

The Effects of Theory-Based Educational Intervention and WhatsApp Follow-up on Papanicolaou Smear Uptake Among Postnatal Women in Malaysia: Randomized Controlled Trial

Zaahirah Mohammad[1*], DrPH; Norliza Ahmad[1*], PhD; Anisah Baharom[1*], PhD

Department of Community Health, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang, Malaysia

[*]all authors contributed equally

Corresponding Author:
Norliza Ahmad, PhD
Department of Community Health
Faculty of Medicine and Health Sciences
Universiti Putra Malaysia
Jalan Universiti 1
Seri Kembangan
Serdang, 43400
Malaysia
Phone: 60 192710577
Email: lizaahmad@upm.edu.my

Abstract

Background: Despite the availability and accessibility of free Papanicolaou (Pap) smear as a screening tool for cervical cancer, the uptake of Pap smear in Malaysia has not changed in the last 15 years. Previous studies have shown that the high uptake of Pap smear reduces the mortality rate of patients with cervical cancer. The low uptake of Pap smear is multifactorial, and the problem could be minimized through the use of mobile technologies. Nevertheless, most intervention studies focused on individual factors, while other important aspects such as mobile technologies, especially WhatsApp, have not been investigated yet.

Objective: This study aims to determine the effects of a theory-based educational intervention and WhatsApp follow-up (Pap smear uptake [PSU] intervention) in improving PSU among postnatal women in Seremban, Negeri Sembilan, Malaysia.

Methods: A 2-arm, parallel single-blind cluster randomized controlled trial was conducted among postpartum women from the Seremban district. Twelve health clinics were randomly assigned to the intervention and control groups. At baseline, both groups received a self-administered questionnaire. The intervention group received standard care and PSU intervention delivered by a researcher. This 2-stage intervention module was developed based on Social Cognitive Theory, where the first stage was conducted face-to-face and the second stage included a WhatsApp follow-up. The control group received standard care. Participants were observed immediately and at 4, 8, and 12 weeks after the intervention. The primary endpoint was PSU, whereas the secondary endpoints were knowledge, attitude, and self-efficacy scores for Pap smear screening self-assessed using a Google Forms questionnaire. A generalized mixed model was used to determine the effectiveness of the intervention. All data were analyzed using IBM SPSS (version 25), and P value of .05 was considered statistically significant.

Results: We analyzed 401 women, of whom 76 (response rate: 325/401, 81%) had withdrawn because of the COVID-19 pandemic, with a total of 162 respondents in the intervention group and 163 respondents in the control group. The proportion of Pap smears at the 12-week follow-up was 67.9% (110/162) in the intervention group versus 39.8% (65/163) in the control group (P<.001). Significant differences between the intervention and control groups were found for Pap smear use ($F_{4,1178}$; P<.001), knowledge scores ($F_{4,1172}$=14.946; P<.001), attitude scores ($F_{4,1172}$=24.417; P<.001), and self-efficacy scores ($F_{1,1172}$=10.432; P<.001).

Conclusions: This study demonstrated that the PSU intervention is effective in increasing the uptake of Pap smear among postnatal women in Seremban district, Malaysia. This intervention module can be tested in other populations of women.

Trial Registration: Thai Clinical Trials Registry TCTR20200205001; https://www.thaiclinicaltrials.org/show/TCTR20200205001

(JMIR Mhealth Uhealth 2022;10(6):e32089) doi:10.2196/32089

KEYWORDS

uterine cervical neoplasms; Papanicolaou test; psychological theory; self-efficacy; social media; health knowledge; attitude; practice; Malaysia

Introduction

Background

Cervical cancer is the fourth most common cancer in women worldwide, after breast, lung, and colon cancer [1]. In 2018, new cases of cervical cancer were estimated at 570,000 [1], with varying incidence and mortality rates in developed and developing countries [2]. Nevertheless, cervical cancer is no longer among the 10 most common cancers in women in developed countries, but is now the second most commonly diagnosed and the third leading cause of cancer-related deaths in women in developing countries [3]. A previous study reported that the global incidence and age-standardized rates of cervical cancer were 15.1 per 100,000 and 13.1 per 100,000, respectively, with a mortality rate of 7.4 per 100,000 [4]. Meanwhile, the incidence rate in developing countries ranged from 10.9 (South Asia) to 43.1 (Africa) [5]. The age-adjusted average varied from 12.8 to more than 20 per 100,000 [6], with a mortality rate between 10 and 30 per 100,000 [5]. In Malaysia, cervical cancer is the third commonest cancer among women with an incidence rate of 6.8 per 100,000, an age-standardized incidence rate of 7.6 per 100,00, and a death rate of 5.6 per 100,000 [1].

The Papanicolaou (Pap) smear screening test can be used for the early diagnosis of cervical cancer and it reduces the risk of cervical cancer–related death by 70% [7]. Nonetheless, at least 80% of women in the recommended population groups must undergo Pap smear screening in order for the program to be effective [8]. Postnatal women represent an important population category that may benefit from opportunistic Pap smears screening given that they are within reach of health services [9]. In Malaysia, Pap smear screening and family planning advice are typically recommended to mothers during their postnatal follow-up [10].

Prior Work

Malaysia's present Pap smear uptake (PSU) is approximately 47.3% [11], which is lower than the intended goal of attaining 70% coverage [7]. This issue of low PSU persists in Malaysia despite the provision of free Pap smear screening tests that are easily accessible in the country. However, currently the test is performed as part of voluntary or opportunistic screening. Previous studies have demonstrated that low Pap smear screening uptake is associated with poor knowledge [12], negative attitude [13], lack of time [14], lack of family support [15], perception of painful procedure [16], lower economic status [17], and embarrassment [16]. Thus, individual and environmental factors play vital roles in determining Pap smear practice among women in Malaysia. In addition, previous intervention studies mainly used the Health Belief Model, which primarily focuses on individual's belief [18-21]. Therefore, other theories need to be explored that focus beyond one's belief. Furthermore, previous studies employed reminder tools such as SMS text messages [22], phone calls [23], and invitation letters [24]. Given that Malaysians are one of the world's largest WhatsApp users with wide coverage [25], using the WhatsApp group as a follow-up platform could be a useful strategy in an intervention program.

Objective

This study aimed to examine the impact of a Social Cognitive Theory (SCT)–based intervention and a WhatsApp follow-up measure, namely PSU intervention, to improve the uptake of Pap smear test among postnatal women in Seremban district, Malaysia. It was hypothesized that postnatal women who received the PSU intervention would have higher uptake 12 weeks after the intervention than women who received standard care.

Methods

Study Design

A 2-arm, parallel, single-blind cluster randomized controlled study was conducted, which comprised an intervention and a control group. The cluster in this study was defined as a health clinic. The intervention arm received the standard care and intervention package, whereas the control arm received only standard routine care. Standard routine care included brief counseling by health care personnel about Pap smear testing and available brochures. The primary outcome was the PSU, whereas the secondary outcomes were participants' knowledge, attitude, and self-efficacy on Pap smear.

Setting and Recruitment

This study was conducted in Seremban, which is the capital city of Negeri Sembilan state in Malaysia with a total population of 620,100 people. Seremban is a developing city that is located about 60 km south of Kuala Lumpur, the capital of Malaysia. It has 12 government health clinics, which are governed by the Seremban Health District Office. This study location was chosen given that the Pap smear screening uptake among women in Seremban district was lower than the national average of 43%, as well as lower than the estimate among postnatal women (39%) [26].

The study population was postnatal women attending Seremban government health clinics. The inclusion criteria were postpartum women who had never participated in Pap smear screening and had a cell phone with WhatsApp installed and internet connection. Meanwhile, the exclusion criteria were postnatal women diagnosed as having cervical cancer, including precancerous stage and postnatal complications, such as postnatal depression, poorly controlled diabetes mellitus, and hypertension. All the aforementioned conditions must have been certified by a medical officer.

Randomization and Allocation Concealment

The 12 health clinics in Seremban district were randomly allocated into the intervention and control groups at a ratio of 1:1. All the postnatal clinics were number coded, whereas simple randomization was performed using Stat Trek software [27].

During participant recruitment, all participants were informed that an intervention was being offered. Therefore, participants were unaware of group assignment throughout the study. Participants were blinded to the fact that awareness of being part of the control group could influence their responses in the questionnaires. These procedures were conducted by a third party who was not involved in this study. The researcher was only aware of the group allocation after the randomization was performed. Systematic random sampling was employed in selecting participants from each postnatal clinic, and those considered eligible and consented were recruited in the study.

Sample Size Calculation

The sample size estimation was based on Lemeshow et al [28] sample size determination in health studies. For hypothesis testing, the formula for 2 population proportions was used to compare the 2 groups. The sample size was calculated using the 2 population proportions formula [29], with a power of 80% to detect a true difference and at a 95% CI. Overall, the sample size was computed based on the uptake of Pap smear test [29], with α of .05 and β of .20, an intraclass correlation coefficient of 0.05, an attrition rate of 20%, and an average cluster size of 10 with a design effect of 1.45. The sample size required after adjusting for the clustered design effect was 394, with 197 participants in each arm.

Intervention

This newly developed intervention module, namely, PSU intervention, used 6 constructs of SCT, comprising cognitive (knowledge), self-efficacy, goal setting, outcome expectation, problem-solving, and reinforcement [30,31]. The module was revised by 2 public health physicians (NA and AB) and 1 family medicine specialist, and the intervention was completed in 2 phases. The first phase was performed via face-to-face and it involved health educational talk and a small group discussion, whereas the second phase entailed a WhatsApp follow-up. This module has been pilot tested among 30 postnatal mothers who are not included in the main study.

This PSU intervention was delivered by ZM who is also a medical doctor. The health educational talk was 15 minutes, which covers the anatomy of female reproductive organs, introduction about cervical cancer, the incidence rate and mortality rate, early diagnostic methods, importance of Pap smear, the positive effects of having Pap smear, and free services available in the government health clinics. This was followed by a 15- to 30-minute small group discussion with approximately 10-15 participants per session. Participants were encouraged to raise any issues or concerns regarding cervical cancer, Pap smear, and any related issues during the face-to-face session. Feeling embarrassed, which was one of the factors that influence PSU, was addressed by using a drape during the screening and this issue was highlighted during the educational talk. It took approximately 30-45 minutes to complete the educational talk and group discussion.

The participants were then recruited in the WhatsApp group for further follow-up and the sessions were conducted weekly for 4 weeks. Allocated time for the WhatsApp group was 1 hour, every Tuesday from 5 to 6 PM. This was the time when the participants were least busy during the week. Nevertheless, participants were also welcome to discuss or ask any questions outside the allocated time. The role of the WhatsApp group was to share information, concerns, and issues; as well as address any misunderstanding on Pap smear and cervical cancer. Besides, it acts as a reminder. The WhatsApp group was made a private group and no other person apart from those recruited by the researcher could access it. No personal information was requested from the WhatsApp group participants and their privacy and confidentiality were protected. Table 1 shows the summary of the contents of educational intervention and WhatsApp follow-up using SCT.

Table 1. Summary of health education intervention contents and WhatsApp follow-up using Social Cognitive Theory.

Number	Social Cognitive Theory constructs	Contents	Method
1	Cognitive (Knowledge)	• Anatomy of women's reproductive system • Information on cervical cancer • Introduction to Pap[a] smear test • Importance of Pap smear test • Misperception of Pap smear test	First phase: • Educational talk • Video on the procedure of a Pap smear test
2	Self-efficacy	• List of situations and scenarios related to Pap smear test • Ways to overcome the issues	First phase: • Group discussion
3	Goal setting	• Setting the goal to undergo a Pap smear test • Setting the goal to adhere to Pap smear practice	First phase: • Educational talk Second phase: • WhatsApp group × 4 weeks
4	Outcome expectation	• Benefits (positive expectation) of Pap smear test • Negative expectations of Pap smear test: embarrassment, discomfort, and minimal pain	First phase: • Educational talk Second phase: • WhatsApp group × 4 weeks
5	Problem-solving	• Problems that might be faced by the participants to undergo Pap smear test	First phase: • Group discussion Second phase: • WhatsApp group × 4 weeks
6	Reinforcement	• Reminders of the importance of Pap smear test and appointment • Reminders of usage of drape during the Pap smear test	Second phase: • WhatsApp group × 4 weeks

[a]Pap: Papanicolaou.

Outcomes

Primary Outcome

The primary outcome was PSU, which was assessed in the intervention and control groups at 4, 8, and 12 weeks after the intervention.

Secondary Outcomes

The secondary outcomes were knowledge, attitude, and self-efficacy assessed immediately and at 4, 8, and 12 weeks after the intervention.

Instrument

A validated self-administered questionnaire that was divided into 6 sections was employed in this study. The first section focused on the participants' sociodemographic characteristics, such as birth date, age, ethnicity, educational level, occupation sector, monthly household income, and marital status. Meanwhile, the second section contained 11 questions on knowledge [18] with the 3 options "yes," "no," or "not sure." The third section also consisted of 11 questions on attitude [18], and participants were instructed to select 1 option from a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The fourth section comprised 14 questions that measured self-efficacy for Pap smear screening, and were evaluated using the Self-Efficacy Scale for Pap Smear Screening Participation (SES-PSSP) Questionnaire [32]. Participants were informed to select only 1 answer from 5 possible options, namely, "definitely," "very likely," "probably," "unlikely," and "definitely not." The fifth section explored the participants' PSU and the option was dichotomous: "yes" or "no." Participants selecting the "yes" option were further instructed to choose the facilities used in conducting their Pap smear tests.

Data Collection

The data collection for this study was conducted from February to December 2020. Data were collected at 5 time points: baseline, immediately after the intervention, and at 4, 8, and 12 weeks after the intervention. Because of the COVID-19 pandemic, attendance at the maternal and child health clinics was severely compromised. Data collection using hard copy self-completed questionnaires was switched to Google Forms for the follow-ups at 4, 8, and 12 weeks after the intervention. The link to the Google Forms was disseminated via the WhatsApp group, and participants had 1 week to complete the questionnaire in the Google Forms.

Data Analysis

All statistical analyses were performed using SPSS (version 25.0; IBM, Inc.). The intention-to-treat principle was used where participants' data were analyzed based on their initially assigned

group. All potential errors were checked prior to data analysis. Descriptive statistics were employed to summarize the data set and the continuous data were assessed for normality. Normally distributed data were presented as mean and SD, whereas median and interquartile ranges (IQR) were used to summarize nonnormally distributed data. Meanwhile, categorical variables were presented in frequencies and percentages.

PSU between the intervention and control groups was compared at each time point using the chi-square test. A generalized linear mixed model was applied to determine the main effects of group, time, and group × time interaction effects for PSU, knowledge, attitude, and self-efficacy between the 2 study groups before and after controlling for covariates. Covariates included were age, ethnicity, education level, and household income. A *P* value of .05 was considered for statistically significant relationships or effects.

Ethical Approval

This study was approved by the Malaysian Medical Research and Ethics Committee, Ministry of Health (Reference number ID: NMRR-19-2589-50455). During data collection, a written and informed consent was obtained from each of the respondents.

Data Sharing

All data relevant to the study are included in the article (also see Multimedia Appendix 1).

Results

Participants' Information

A total of 401 eligible participants (intervention group: n=201, control group: n=200) were recruited in this study. The overall response rate was 81% (325/401) at 12 weeks after the intervention. Figure 1 illustrates the CONSORT-eHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth) flowchart [33] of the study.

No statistically significant difference was detected between the intervention and control groups at baseline for covariates and outcomes measures (Table 2). A total of 0.26% of the data were missing completely at random (χ^2_2=0.867; *P*=.64). Among those who could not be followed up, most were between 26 and 30 years of age, of Malay descent, had a tertiary education, and were in the M40 (RM 2802-RM 5865 [US $634.15-US $1327.37]) income category.

Most participants in both groups were between the ages 26 and 30 years, of Malay ethnicity, married, government servants, and attained tertiary educational level. Household income was categorized into 3 groups: below 40% (B40), middle 40% (M40), and top 20% (T20) of the Negeri Sembilan household income [34]. Most participants were in the M40 category, and 86.5% (347/401) of the participants had Malay ethnicity, followed by Chinese, Indian, and others.

Figure 1. CONSORT-eHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and onLine TeleHealth) flowchart [33].

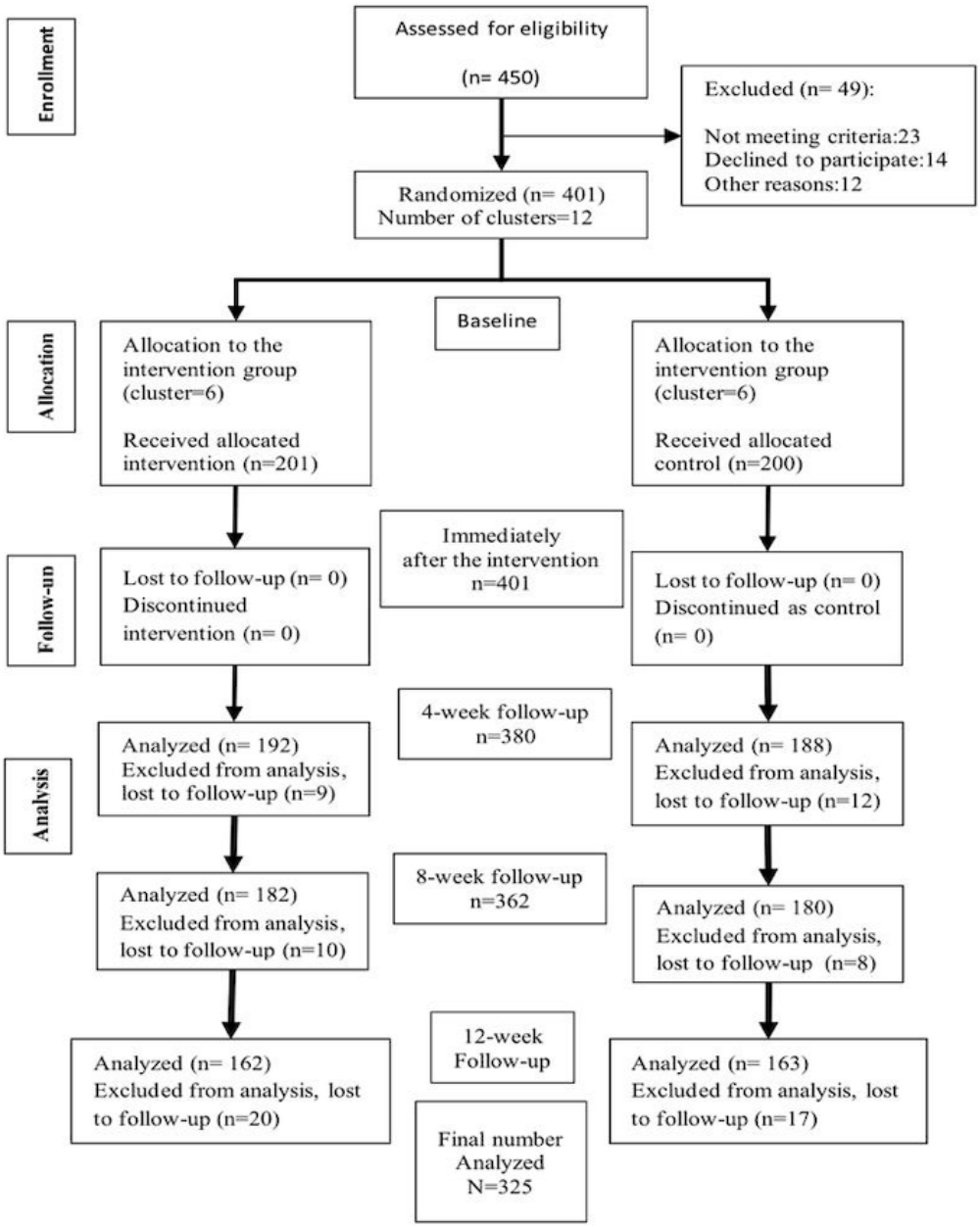


Table 2. Participants' baseline characteristics.

Variables	Intervention (n=201), n (%)	Control (n=200), n (%)	Difference between the conditions	
			Statistical test (*df*)	*P* value
Age (years)			7.892[a] (1)	.25
20-25	24 (11.9)	19 (9.5)		
26-30	78 (38.8)	93 (46.5)		
31-35	53 (26.4)	48 (24.0)		
>36	46 (22.9)	40 (20.0)		
Ethnicity			8.349[a] (1)	.54
Malay	178 (88.6)	169 (84.5)		
Chinese	15 (7.5)	20 (10.0)		
Indian	8 (4.0)	11 (5.5)		
Marital status			—[b]	.43
Married	197 (98.0)	196 (98.0)		
Single mother	4 (2.0)	4 (2.0)		
Level of education			0.871[a] (1)	.16
Secondary	76 (37.8)	68 (34.0)		
Tertiary	125 (62.2)	132 (66.0)		
Occupation sector			0.768[a] (1)	.18
None	43 (21.4)	39 (19.5)		
Government	126 (62.7)	128 (64.0)		
Private	20 (10.0)	24 (12.0)		
Self-employed	12 (6.0)	9 (4.5)		
Household income[c]			0.768[a] (1)	.25
B40 (<RM[d,e] 2801)	34 (16.9)	32 (16.0)		
M40 (RM 2802-5865)	122 (60.7)	114 (57.0)		
T20 (≥RM 5866)	45 (22.4)	54 (27.0)		
Knowledge scores, median (IQR)	8 (3)	6 (4)	1221[f]	.84
Attitude scores, median (IQR)	28 (5)	26 (5)	1090.5[f]	.27
Self-efficacy scores, median (IQR)	38 (8)	36 (7)	1119[f]	.68

[a]Chi-square test.

[b]Fisher exact test.

[c]Household income was categorized into the following based on the Department of Statistics Malaysia: B40, M40, and T20 (specific for Negeri Sembilan).

[d]RM: Malaysian Ringgit.

[e]1 RM=1 US $0.23.

[f]Mann-Whitney *U* test.

Primary Outcome: Pap Smear Test Uptake

The participants' PSU at all the time points is presented in Table 3. The results showed that significantly (P<.001) more respondents in the intervention group than in the control group had a Pap smear performed before and after controlling for covariates. There was a significant difference in the intervention group at baseline, 4 weeks after the intervention, 8 weeks after the intervention, and 12 weeks after the intervention with $F_{4,1178}$=3.222 and P<.001.

Table 3. Proportion of Papanicolaou smear test uptake among participants at each time point.

Variable	Immediately after the intervention	4 weeks after the intervention	8 weeks after the intervention	12 weeks after the intervention
Intervention, n/N (%)	0/201 (0)	54/193 (27.9)	86/183 (46.9)	110/162 (67.9)
Control, n/N (%)	1/200 (0.5)	24/185 (12.9)	56/181 (30.9)	65/163 (39.8)
Chi-square test	256	54	89	125
df	1	1	1	1
P value	.66	.04[a]	.02[a]	<.001[a]

[a]Statistically significant.

Secondary Outcomes: Educational Intervention and WhatsApp Follow-up Outcome

Table 4 presents the generalized linear mixed model results for the total knowledge, attitude, and self-efficacy scores and participants' intention to adhere to Pap smear practice after controlling for the covariates. The results indicated that significantly more respondents in the intervention group than in the control group had increased their total scores of knowledge ($F_{1,1171}$=14.946, P<.001), total scores of attitude ($F_{1,1171}$=14.946; P<.001), and total scores for self-efficacy ($F_{1,1171}$=10.432, P<.001).

Table 4. Effects of educational intervention and WhatsApp follow-up on knowledge, attitude, self-efficacy scores, and intention to adhere to Papanicolaou smear practice among postnatal mothers.

Variables and parameters	*F*	*df1*	*df2*	*P* value[a]
Knowledge scores				
Group	1273	1	1172	<.001[b]
Time	11.658	1	1172	<.001[b]
Group × time	14.946	4	1172	<.001[b]
Attitude scores				
Group	458	1	1172	<.001[b]
Time	35.12	4	1172	<.001[b]
Group × time	24.417	4	1172	<.001[b]
Self-efficacy scores				
Group	292.038	1	1172	<.001[b]
Time	13.254	4	1172	<.001[b]
Group × time	10.432	4	1172	<.001[b]

[a]Using a generalized linear mixed model adjusted for participants' age, ethnicity, education level, and household income.

[b]Statistically significant.

Discussion

Principal Findings

The aim of this study was to evaluate the effects of a PSU intervention based on SCT and using WhatsApp follow-up to improve PSU, knowledge, attitude, and self-efficacy among postpartum women; 12 weeks after the intervention, the intervention group demonstrated a significant increase in PSU. The intervention group also recorded significantly higher knowledge, attitude, and self-efficacy compared with the control group.

Behavioral changes among participants in the intervention group could be attributed to the SCT constructs employed in this study. Given that the health intervention provided a clear picture of cervical cancer and Pap smear test, the participants might have been influenced to have specific goals to undergo the Pap smear test and adhere to Pap smear practice. All the barriers that might arise were discussed comprehensively during the group discussion, which assisted participants in problem-solving and improved their self-efficacy and their PSU. Furthermore, the expected outcome was highlighted during the educational talk and WhatsApp follow-up. Concerns related to the Pap smear test, such as embarrassment, slight discomfort, and minimal pain, were shared with the participants. This information might have motivated the participants to be physically and mentally ready to undergo the Pap smear screening test.

The weekly reminders through the WhatsApp group follow-up reinforced the importance of Pap smear screening and assisted

the participants in booking Pap smear screening appointments. Furthermore, the participants were reminded of the usage of drapes and privacy policy during Pap smear test to reduce the feeling of embarrassment. These procedures might have contributed to their positive attitude toward the Pap smear screening test. Self-efficacy is crucial as a determinant of a woman's decision to uptake the Pap smear. Poor self-efficacy was reported to be influenced by spouses and family members in several developing countries [30]. The SCT constructs used in our study might have contributed to the enhanced self-efficacy observed among participants as depicted in the postintervention self-efficacy scores, which might have encouraged them to undergo the Pap smear screening test.

Comparison With Prior Literature

The study's finding on PSU is consistent with a previous randomized control study conducted using SCT in which 70% of participants in the intervention group underwent a Pap smear test [31]. In their study, some of the utilized SCT constructs were cognitive (knowledge), goal setting, and self-efficacy. Nevertheless, the study by Wang et al [31] employed only a face-to-face method, whereas this study utilized the WhatsApp mobile app as an additional follow-up method. Another disparity is the relatively shorter duration in this study compared with 12 months' follow-up in the previous study [31]. The use of WhatsApp follow-up might assist in reducing the follow-up duration while achieving similar Pap smear screening uptake results. Given the wide acceptance of the WhatsApp platform, this approach could be client-friendly as a follow-up modality that could serve as a reminder and to resolve issues [35].

The improved knowledge scores of participants in this study were consistent with the findings from other interventional studies [36-38]. For instance, participants' knowledge scores on the importance of Pap smear were significantly impacted following an interactive session in a randomized controlled trial conducted in Korea, which focused on the anatomy of female genitalia and cervical cancer [38]. Another reason for the improved knowledge scores could be due to the participants' education level. Most participants in this study attained tertiary educational level and their motivation to seek knowledge was higher compared with those with secondary and primary educational levels [39].

Strengths and Limitations

This is a cluster randomized controlled trial with a good response rate despite the COVID-19 pandemic that occurred during the data collection. A few crucial constructs of SCT were employed as the educational intervention in this study, which was delivered through a face-to-face session and WhatsApp follow-up. This study is among the few local studies investigating the effects of an intervention on PSU, knowledge, attitude, self-efficacy, and intention to adhere to Pap smear practice. To date, this is the first study to employ an educational intervention and a subsequent follow-up technique and reminders for PSU using WhatsApp.

Some limitations of this study include self-reported questionnaires, which may lead to either underreporting or overreporting of results, especially regarding participants' self-efficacy. This study was conducted in government health clinics, which might have contributed to the low participation of other ethnicities. Specifically, 86.5% (347/401) of the participants were of Malay ethnicity, which was different from the Malaysian demographics pattern of 68.6% [26]. Other women populations were not included in this study as the inclusion criteria entailed postnatal women aged 20-49 years old. Hence, the findings may be different among other nonpostnatal women.

With limited human resources, replicating this intervention might be difficult as it will be an additional burden to the staff at health clinic levels. The WhatsApp group follow-up might be feasible for some health care facilities; however, other centers might find this approach time-consuming and laborious. Factors such as suitability, timing, and human resources need to be considered before implementing an intervention at the clinic level.

Conclusion

This study suggests that SCT-based health education intervention and WhatsApp group follow-up are effective to improve the PSU among postnatal women, as well as their knowledge, attitude, and self-efficacy. This intervention can be evaluated in other populations that are more representative of Malaysian women and can also be used as baseline data for other intervention studies.

Acknowledgments

We thank the Director-General of Health Malaysia for authorizing the publication of this article. We are grateful to the Seremban District Health Office and all participating health clinics and the participants for their amazing support and co-operation. We are also grateful to Azline Abdilah and Aneesa Abdul Rashid for their involvement in the intervention study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2784 KB - mhealth_v10i6e32089_app1.pdf]

References

1. World Health Organization. Be he@lthy, be mobile: a handbook on how to implement mCervicalCancer. Geneva, Switzerland: World Health Organization; 2017.
2. Vaccarella S, Laversanne M, Ferlay J, Bray F. Cervical cancer in Africa, Latin America and the Caribbean and Asia: Regional inequalities and changing trends. Int J Cancer 2017 Nov 15;141(10):1997-2001 [FREE Full text] [doi: 10.1002/ijc.30901] [Medline: 28734013]
3. Zhang S, Xu H, Zhang L, Qiao Y. Cervical cancer: Epidemiology, risk factors and screening. Chin J Cancer Res 2020 Dec 31;32(6):720-728 [FREE Full text] [doi: 10.21147/j.issn.1000-9604.2020.06.05] [Medline: 33446995]
4. Arbyn M, Weiderpass E, Bruni L, de Sanjosé S, Saraiya M, Ferlay J, et al. Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. The Lancet Global Health 2020 Feb;8(2):e191-e203. [doi: 10.1016/s2214-109x(19)30482-6]
5. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Nov;68(6):394-424 [FREE Full text] [doi: 10.3322/caac.21492] [Medline: 30207593]
6. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin 2015 Mar 04;65(2):87-108 [FREE Full text] [doi: 10.3322/caac.21262] [Medline: 25651787]
7. Wingo PA, Cardinez CJ, Landis SH, Greenlee RT, Ries LAG, Anderson RN, et al. Long-term trends in cancer mortality in the United States, 1930-1998. Cancer 2003 Jun 15;97(12 Suppl):3133-3275 [FREE Full text] [doi: 10.1002/cncr.11380] [Medline: 12784323]
8. Anttila A, Ronco G, Clifford G, Bray F, Hakama M, Arbyn M, et al. Cervical cancer screening programmes and policies in 18 European countries. Br J Cancer 2004 Aug 31;91(5):935-941 [FREE Full text] [doi: 10.1038/sj.bjc.6602069] [Medline: 15280916]
9. Ago BU, Etokidem A, Ebughe G. Prevalence of Abnormal Cervical Cytology among Postnatal Clinic Attendees at the University of Calabar Teaching Hospital, Nigeria. OALib 2016;03(09):1-14. [doi: 10.4236/oalib.1102992]
10. Yunus NA, Mohamed Yusoff H, Draman N. Non-Adherence to recommended Pap smear screening guidelines and its associated factors among women attending health clinic in Malaysia. Malays Fam Physician 2018;13(1):10-17 [FREE Full text] [Medline: 29796205]
11. Summary of Malaysian National Cancer Registry Report 2007-2011. Ministry of Health of Malaysia. 2015. URL: https://www.moh.gov.my/moh/resources/Penerbitan/Rujukan/NCD/Kanser/Summary_of_Malaysian_National_Cancer_Registry_Report_2007-2011.pdf [accessed 2022-06-09]
12. Alwahaibi N, Alsalami W, Alramadhani N, Alzaabi A. Factors Influencing Knowledge and Practice Regarding Cervical Cancer and Pap smear Testing among Omani Women. Asian Pac J Cancer Prev 2018 Dec 01;19(12):3367-3374. [doi: 10.31557/apjcp.2018.19.12.3367]
13. Tapera R, Manyala E, Erick P, Maswabi TM, Tumoyagae T, Letsholo B, et al. Knowledge and Attitudes towards Cervical Cancer Screening amongst University of Botswana Female Students. Asian Pac J Cancer Prev 2017 Sep 27;18(9):2445-2450 [FREE Full text] [doi: 10.22034/APJCP.2017.18.9.2445] [Medline: 28952274]
14. Al-Amro S, Gharaibeh M, Oweis A. Factors Associated with Cervical Cancer Screening Uptake: Implications for the Health of Women in Jordan. Infect Dis Obstet Gynecol 2020;2020:9690473 [FREE Full text] [doi: 10.1155/2020/9690473] [Medline: 32256030]
15. George T J. Factors influencing utilization of cervical cancer screening services among women – A cross sectional survey. Clinical Epidemiology and Global Health 2021 Jul;11:100752. [doi: 10.1016/j.cegh.2021.100752]
16. Black E, Hyslop F, Richmond R. Barriers and facilitators to uptake of cervical cancer screening among women in Uganda: a systematic review. BMC Womens Health 2019 Aug 09;19(1):108 [FREE Full text] [doi: 10.1186/s12905-019-0809-z] [Medline: 31399092]
17. Broberg G, Wang J, Östberg AL, Adolfsson A, Nemes S, Sparén P, et al. Socio-economic and demographic determinants affecting participation in the Swedish cervical screening program: A population-based case-control study. PLoS One 2018 Jan 10;13(1):e0190171 [FREE Full text] [doi: 10.1371/journal.pone.0190171] [Medline: 29320536]
18. Rezaie-Chamani S, Mohammad-Alizadeh CS, Kamalifard M. Knowledge, attitudes and practice about Pap smear among women reffering to a public hospital. J Fam Reprod Heal 2012;6(4):177-182 [FREE Full text]
19. O'Brien MJ, Halbert CH, Bixby R, Pimentel S, Shea JA. Community health worker intervention to decrease cervical cancer disparities in Hispanic women. J Gen Intern Med 2010 Nov 7;25(11):1186-1192 [FREE Full text] [doi: 10.1007/s11606-010-1434-6] [Medline: 20607434]
20. Romli R, Shahabudin S, Saddki N, Mokhtar N. Effectiveness of a Health Education Program to Improve Knowledge and Attitude Towards Cervical Cancer and Pap Smear: A Controlled Community Trial in Malaysia. Asian Pac J Cancer Prev 2020 Mar 01;21(3):853-859. [doi: 10.31557/apjcp.2020.21.3.853]
21. Broberg G, Jonasson JM, Ellis J, Gyrd-Hansen D, Anjemark B, Glantz A, et al. Increasing participation in cervical cancer screening: telephone contact with long-term non-attendees in Sweden. Results from RACOMIP, a randomized controlled trial. Int J Cancer 2013 Jul 10;133(1):164-171 [FREE Full text] [doi: 10.1002/ijc.27985] [Medline: 23233356]
22. Linde DS, Andersen MS, Mwaiselage J, Manongi R, Kjaer SK, Rasch V. Effectiveness of One-Way Text Messaging on Attendance to Follow-Up Cervical Cancer Screening Among Human Papillomavirus-Positive Tanzanian Women

(Connected2Care): Parallel-Group Randomized Controlled Trial. J Med Internet Res 2020 Apr 02;22(4):e15863 [FREE Full text] [doi: 10.2196/15863] [Medline: 32238335]
23. Abdul Rashid RM, Mohamed M, Hamid ZA, Dahlui M. Is the phone call the most effective method for recall in cervical cancer screening?--results from a randomised control trial. Asian Pac J Cancer Prev 2013 Oct 30;14(10):5901-5904 [FREE Full text] [doi: 10.7314/apjcp.2013.14.10.5901] [Medline: 24289597]
24. Tavasoli SM, Pefoyo AJK, Hader J, Lee A, Kupets R. Impact of invitation and reminder letters on cervical cancer screening participation rates in an organized screening program. Prev Med 2016 Jul;88:230-236. [doi: 10.1016/j.ypmed.2016.04.019] [Medline: 27143497]
25. Bhat R. WhatsApp Statistics for 2022 - All You Need to Know. Verloop. 2022. URL: https://verloop.io/blog/whatsapp-statistics/#whatsapp-gender-demographics [accessed 2022-06-09]
26. Ministry of Health of Malaysia. Annual Report Ministry of Health Malaysia 2017. Ministry of Health of Malaysia. 2017. URL: https://www.moh.gov.my/moh/resources/Penerbitan/Penerbitan%20Utama/Annual%20Report%20MoH%202017.pdf [accessed 2022-06-09]
27. Alford SH, Leadbetter S, Rodriguez JL, Hawkins NA, Scholl LE, Peipins LA. Cancer screening among a population-based sample of insured women. Prev Med Rep 2015;2:15-20 [FREE Full text] [doi: 10.1016/j.pmedr.2014.11.004] [Medline: 26844046]
28. Lemeshow S, Hosmer D, Klar J, Lwanga S, Organization W. Adequacy of Sample Size in Health Studies. Chichester, UK: Wiley; 1990.
29. Donner A, Birkett N, Buck C. Randomization by cluster. Sample size requirements and analysis. Am J Epidemiol 1981 Dec;114(6):906-914. [doi: 10.1093/oxfordjournals.aje.a113261] [Medline: 7315838]
30. Byrd TL, Wilson KM, Smith JL, Coronado G, Vernon SW, Fernandez-Esquer ME, et al. AMIGAS: a multicity, multicomponent cervical cancer prevention trial among Mexican American women. Cancer 2013 Apr 01;119(7):1365-1372 [FREE Full text] [doi: 10.1002/cncr.27926] [Medline: 23280399]
31. Wang X, Fang C, Tan Y, Liu A, Ma GX. Evidence-based intervention to reduce access barriers to cervical cancer screening among underserved Chinese American women. J Womens Health (Larchmt) 2010 Mar;19(3):463-469 [FREE Full text] [doi: 10.1089/jwh.2009.1422] [Medline: 20156089]
32. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007 Jan;60(1):34-42. [doi: 10.1016/j.jclinepi.2006.03.012] [Medline: 17161752]
33. 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]
34. Kenyataan Media Bagi Laporan Perangkaan Demografi Malaysia: Suku Tahun Pertama 2021. Department of Statistics Malaysia. 2021. URL: https://www.dosm.gov.my/v1/uploads/files/5_Gallery/2_Media/4_Stats%40media/4-Press_Statement/2021/20210412-Kenyataan_Media-Penggunaan_dan_Capaian_ICT_oleh_Individu_dan_Isi_Rumah_2020.pdf [accessed 2022-06-09]
35. 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]
36. Soto M, Martinez-Gutierrez J, Momany M, Capurro D, Ciampi Spode F, Cea E, et al. Preferences of Underserved Chilean Women on a Mobile Technology Intervention for Cervical Cancer Screening: Qualitative Study. JMIR Mhealth Uhealth 2018 Nov 20;6(11):e196 [FREE Full text] [doi: 10.2196/mhealth.9494] [Medline: 30459141]
37. Abdullah F, O'Rorke M, Murray L, Su T. Evaluation of a worksite cervical screening initiative to increase Pap smear uptake in Malaysia: a cluster randomized controlled trial. Biomed Res Int 2013;2013:572126 [FREE Full text] [doi: 10.1155/2013/572126] [Medline: 24073411]
38. Park S, Chang S, Chung C. Effects of a cognition-emotion focused program to increase public participation in Papanicolaou smear screening. Public Health Nurs 2005 Jul;22(4):289-298. [doi: 10.1111/j.0737-1209.2005.220404.x] [Medline: 16150010]
39. Lawrence EM. Why Do College Graduates Behave More Healthfully than Those Who Are Less Educated? J Health Soc Behav 2017 Jun 28;58(3):291-306 [FREE Full text] [doi: 10.1177/0022146517715671] [Medline: 28845056]

Abbreviations

CONSORT-eHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth
Pap: Papanicolaou
PSU: Pap smear uptake
SCT: Social Cognitive Theory
SES-PSSP: Self-Efficacy Scale for Pap Smear Screening Participation

Edited by L Buis; submitted 14.07.21; peer-reviewed by Y Motoki, A Roundtree; comments to author 22.12.21; revised version received 30.03.22; accepted 18.04.22; published 27.06.22.

Please cite as:
Mohammad Z, Ahmad N, Baharom A
The Effects of Theory-Based Educational Intervention and WhatsApp Follow-up on Papanicolaou Smear Uptake Among Postnatal Women in Malaysia: Randomized Controlled Trial
JMIR Mhealth Uhealth 2022;10(6):e32089
URL: https://mhealth.jmir.org/2022/6/e32089
doi:10.2196/32089
PMID:35759319

Original Paper

Real-world Benefits of Diabetes Management App Use and Self-monitoring of Blood Glucose on Glycemic Control: Retrospective Analyses

Ya-Ting Chang[1], BSc; Yu-Zhen Tu[1], MSc; Hung-Yi Chiou[1], PhD; Ken Lai[1], MSc; Neng Chun Yu[2], MD

[1]H2 Inc, Taipei, Taiwan

[2]Neng Chun Diabetes Clinic, Medical & Health in Lotung, Yilan, Taiwan

Corresponding Author:
Ken Lai, MSc
H2 Inc
4F, No 32, Aly 18, Ln 478
Ruiguang Rd, Neihu Dist
Taipei, 114063
Taiwan
Phone: 886 287976661 ext 105
Email: klai@health2sync.com

Abstract

Background: Among self-care measures, the self-monitoring of blood glucose (SMBG) is a critical component for checking blood glucose levels. In addition, there is growing evidence suggesting that digital technologies are being adopted as an additional method for health care systems to increase patient contact. However, for patients with non–insulin-treated diabetes mellitus type 2 (DMT2), the value of SMBG was inconsistent among studies, and the evidence for digital technologies from real-world clinical practice is still limited.

Objective: Our study aimed to assess patients with non–insulin-treated DMT2 who were receiving care from a single clinic and analyze whether the use of a diabetes management app and SMBG behavior would affect glycemic control in a real-world clinical setting.

Methods: We collaborated with a large clinic focused on diabetes care in Taiwan that had been using the Health2Sync mobile app and web-based Patient Management Platform to collect the data. The patients were divided into 2 groups (app-engaged-user group and only-data-uploader group) according to different activities in the app, and blood glucose was recorded every month from 1 to 6 months after registration in the app. A sample of 420 patients was included in the analysis, and a linear mixed model was built to investigate which factors affected the patients' blood glucose percentage change.

Results: Using the mixed model coefficient estimates, we found that the percentage change was significantly negative when the only-data-uploader group was set as the baseline (t=–3.873, df=1.81 × 10^4; P<.001 for the patients of the app-engaged-user group). We found that for patients with shorter diabetes duration, their blood glucose decreased more than patients with longer diabetes duration (t=2.823, df=1.71 × 10^4; P=.005 for the number of years of diabetes duration). In addition, we found that for younger patients, their blood glucose decreased more than older patients (t=2.652, df=1.71 × 10^4; P=.008 for the age of the patients). Furthermore, the patients with an education level of junior high school or lower saw a significantly greater decrease in blood glucose percentage change than the patients with an education level of senior high school or higher (t=4.996, df=1.72 × 10^4; P<.001 for the patients with an education level of senior high school or higher). We also found that the count of blood glucose measured enlarged the decrease along the interaction months (t=–8.266, df=1.97 × 10^4; P<.001 for the nth month × the count of blood glucose in the nth month). Lastly, the gender of the patients did not significantly affect the percentage change (t=0.534, df=1.74 × 10^4; P=.59 for female patients).

Conclusions: Our analysis showed the following: the blood glucose percentage change of the patients in the app-engaged-user group dropped more than that in the only-data-uploader group; shorter diabetes duration is associated with a steeper decrease in the patients' blood glucose percentage change; the percentage decrease in blood glucose change in younger patients is greater than older patients; the blood glucose percentage change of the patients with an education level of junior high school or lower dropped more than those with an education level of senior high school or higher; and the more frequently the patients test SMBG

each month, the greater the decrease in the patients' blood glucose percentage. Further studies can be performed to consider the differences in daily behaviors such as exercise and diet across the patients and whether these factors could have vital effects on glycemic control.

(JMIR Mhealth Uhealth 2022;10(6):e31764) doi:10.2196/31764

KEYWORDS

diabetes care; digital intervention; mobile app; real-world data; glycemic control; mobile health; digital therapeutics; diabetes; therapy; app

Introduction

Many studies have shown that diabetes mellitus not only results in specific complications but also leads to increased risks of cardiovascular disease and cancer [1-3]. Although it might have adverse outcomes, diabetes is now considered a chronic disorder that can usually be controlled with appropriate treatment, lifestyle management, and self-care measures to keep blood glucose in the target range [4,5].

Among self-care measures, the self-monitoring of blood glucose (SMBG) is a critical component for checking blood glucose levels [5-7]. Several studies have provided evidence that SMBG has notable benefits on glycemic control, and a recent meta-analysis showed that SMBG has beneficial effects on glucose control in both the short- and long-term [8-10]. Specifically, previous research articles have shown that SMBG is helpful for patients with diabetes mellitus type 1 in controlling blood sugar level. Furthermore, one randomized controlled trial (RCT) recruited patients with diabetes mellitus type 2 (DMT2) and observed them for at least 12 months, and the results suggested that SMBG improves diabetes control [11]. However, for patients with non–insulin-treated DMT2, the value of SMBG was inconsistent among the studies [11-15]. The difference may be due to the different research designs or targets of diabetes type in the studies.

To address the limitations of previous studies, we focused on patients with non–insulin-treated DMT2 who were receiving care from a single clinic and investigated the relationship between the frequency of SMBG and the patients' glucose levels. In addition, there is growing evidence suggesting that digital technologies are being adopted as an additional method for health care systems to increase patient contact and enhance the effect of conventional care practices for diabetes patients [16,17]. We also focused on the patients who used a diabetes management app with self-care measures during the observation period. Therefore, the objective of this study was to analyze whether the app and SMBG affected glycemic control.

Methods

Data Collection

The Health2Sync mobile app and web-based Patient Management Platform were used to collect the data. These products were described in our previous study [16]. We collaborated with a large clinic focused on diabetes care in Taiwan that had been using these products. All the patients analyzed in this study belonged to the same diabetes clinic and received the clinic's standard care. During a patient's visit, the clinic's health care professionals (HCPs) helped the patient register an account in the app and collected the patient's demographic data, including gender, age, diabetes type, diabetes duration, and education level. After registration, smartphone-proficient users who were willing to use a digital management solution would start to use the Health2Sync mobile app; otherwise, patients would let HCPs sync their SMBG records from the blood glucose meters to their accounts during subsequent visits.

To assess the effects of SMBG and digital intervention from the Health2Sync mobile app, the patients' SMBG records were averaged on a monthly basis, with the patients' average in the first week after app registration designated as the baseline. Since each individual patient had different SMBG habits, only fasting blood glucose (FBG) records were included in the analyses for comparison. As the baseline blood glucose level of each patient was different, the maximum magnitude of the blood glucose increase or decrease could also be different, so we used the blood glucose percentage change instead of the blood glucose value change to assess the glycemic status improvement across the groups. The formula for that percentage change for each patient was *(mean of blood glucose value of each month – baseline blood glucose level) / baseline blood glucose level*.

Patient Inclusion

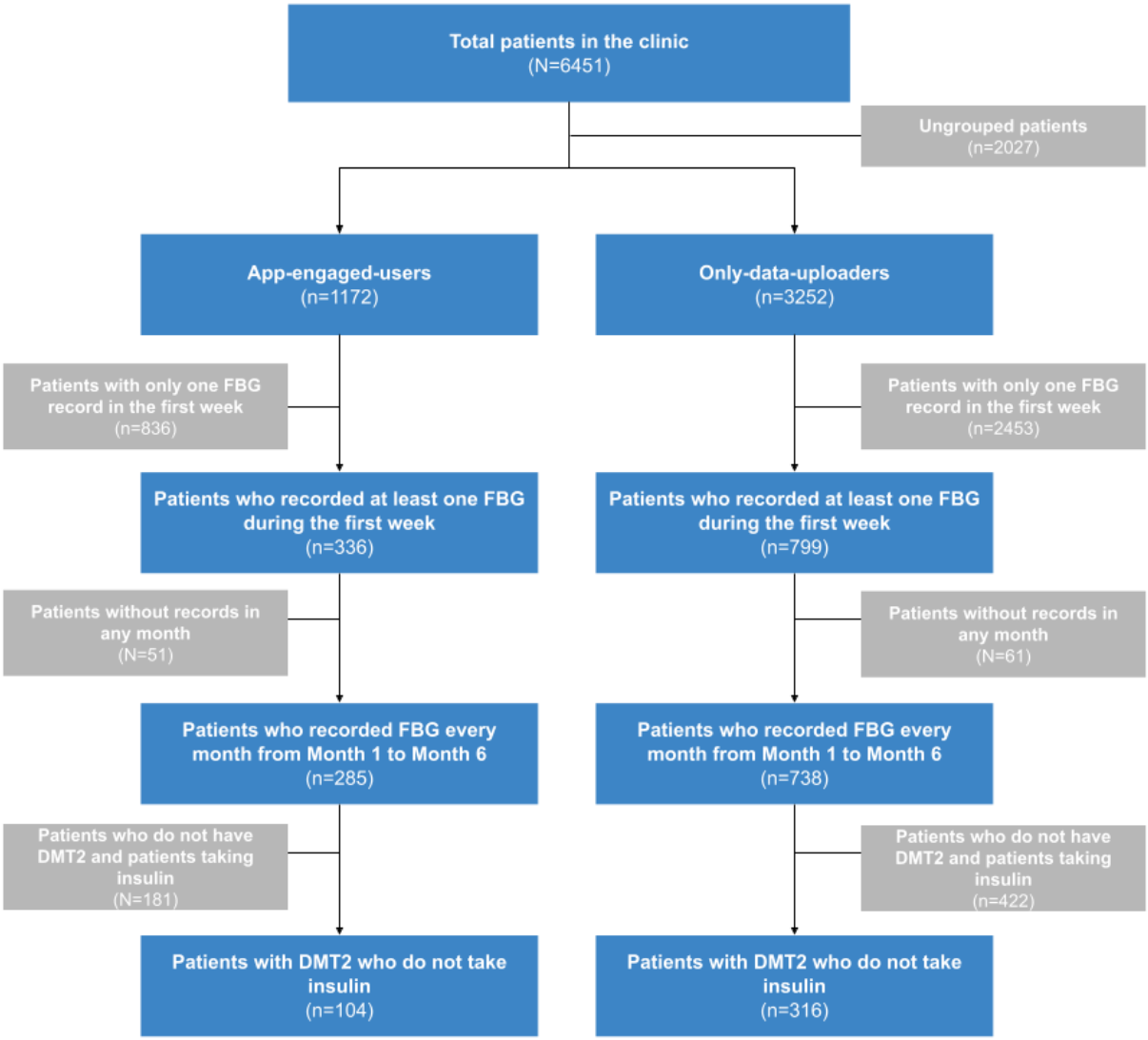
The clinic had 6451 app-registered patients. To separate patients with different activities in the app, we defined 2 groups of users based on their frequency of using the Health2Sync mobile app. After registering for the app, the patients who used the app at least once a week on average were labeled as "app-engaged-users," and those who used the app at most once a month on average were labeled as "only-data-uploaders," as we believed that their SMBG were mainly uploaded by HCPs and they seldom used the app at home. The rest of the patients were excluded. At this stage, we excluded 2027 patients, leaving 1172 patients in the app-engaged-user group and 3252 patients in the only-data-uploader group.

To calculate the patients' blood glucose level change, we took the mean value of each patient's FBG value recorded in the first week after app registration. Patients with only one FBG record in the first week were excluded as we believed this value was unrepresentative of the blood glucose level in the whole week. At this stage, we excluded 836 patients from the app-engaged-user group, with 336 remaining, and 2453 patients from the only-data-uploader group, with 799 remaining. The patients who had no record in any month from 1 to 6 months after app registration were also excluded, because a complete data set would be needed for later analyses, where blood glucose

level is the dependent variable in modeling. At this stage, we excluded 51 and 61 patients from app-engaged-user and only-data-uploader groups, respectively. Finally, to eliminate the impact from differences due to diabetes type and medication, only DMT2 patients who do not take insulin treatments were included for the analyses. Eventually, we had 104 and 316 patients in the app-engaged-user and only-data-uploader groups, respectively. Figure 1 shows the inclusion flow chart described above.

In addition, we analyzed the blood glucose meters used by the patients. We were able to collect the blood glucose meter information for 320 patients. Among those, 302 patients used blood glucose meters that are compliant with the requirements for blood-glucose monitoring systems for self-testing in managing diabetes mellitus (ISO 15197:2013) [18], so we believe our study is based on accurate SMBG data (Multimedia Appendix 1).

Figure 1. Inclusion flow chart of this study. DMT2: diabetes mellitus type 2; FBG: fasting blood glucose.



Analysis

Software and Model Used

We used R statistical software (version 3.6.1; R Foundation for Statistical Computing) [19] for all the statistical analyses, including *t* test (2-tailed), Pearson chi-square test, and one-way ANOVA. We also used the linear mixed model of the lme4 package for R (version 2015; Bates et al [20]).

Patient Characteristics

We used one-way ANOVA and Pearson chi-squared test for continuous and categorical variables, respectively, to check the homogeneity of the demographics across the 2 groups. In addition, we used one-way ANOVA to test whether the initial blood glucose levels of the 2 groups are different.

Statistical Modeling and Analysis

We started by creating a model (our original model) that included the key factors we believed would affect a patient's blood glucose percentage change. We also wanted to include time (the nth month) as a factor for analysis, so a linear mixed model was used to analyze our original model. However, there were a few factors that were significantly different between the 2 patient groups. To confirm whether our model should include the interaction of these factors and the patient groups, we built a basic model that only included the patient groups and factors that were significantly different to check whether these factors

had an effect on the blood glucose percentage change. Subsequently, for each of the significantly different factors, we built a new model based on the basic model that included the interaction of that individual factor with the groups. Finally, we built a new model that added all the factor-group interactions to the basic model to confirm whether the interactions of these factors and the groups have an effect on the blood glucose percentage change. We used the *P* value to determine whether these interaction factors should be added back into the original model [21].

Ethical Considerations

Institutional Review Board approval was not sought for this study as it is based on retrospective analysis, and patients can freely choose whether or not to use the Health2Sync mobile app. The patients in both the app-engaged-user and only-data-uploader groups agreed to Health2Sync's Privacy Policy before registering an account, giving H2 Inc the right to analyze their data for research purposes.

Results

Table 1 presents the patients' demographic characteristics stratified by the 2 groups. There were no significant differences in gender (X^2_1=0; *P*>.99), diabetes duration (t_{418}=–0.69; *P*=.49), and the baseline blood glucose level (t_{418}=–0.58; *P*=.56) between the 2 groups. However, significant differences were found in age (t_{418}=–6.66; *P*<.001) and education level (X^2_1=45.44; *P*<.001).

Table 1. Patient characteristics analyses.

Characteristic	All patients (n=420)	App-engaged-user group (n=104)	Only-data-uploader group (n=316)	*P* value
Age (years), mean (SD)	59.28 (11.29)	52.7 (12.06)	61.44 (10.16)	<.001
Gender, n (%)				>.99
Male	219 (52.1)	54 (51.9)	165 (52.2)	
Female	201 (47.9)	50 (48.1)	151 (47.8)	
Diabetes duration (years), mean (SD)	9.23 (7.95)	8.74 (8.46)	9.39 (7.78)	.49
Education level, n (%)				<.001
Junior high school or lower	181 (46)	19 (20)	163 (54)	
Senior high school or higher	212 (54)	75 (80)	137 (46)	
Initial blood glucose level (mg/dL), mean (SD)	135.62 (31.34)	134.16 (28.29)	136.10 (32.31)	.56

Linear mixed modeling was used to estimate the effects from factors that could affect the patients' blood glucose percentage change. In addition, due to the above analysis of patient characteristics, we know that there are significant differences in the patients' age and education level between the patient groups. Therefore, we have to confirm whether the interactions of these factors and the groups should be put into the original model.

First, we built a basic model to check whether the patients' group, age, and education level have an effect on the blood glucose percentage change. This basic model only included the patients' group, age, and education level; these factors exhibited significant effects on the blood glucose percentage change (*P*<.001 for the patients' age; *P*<.001 for the patients' education level; *P*<.001 for the patients' group). Second, we wanted to examine whether an interaction effect of the patients' age and group has an effect on the blood glucose percentage change. We built a second model that included the same factors as the basic model, but also added an interaction effect of the patients' age and group. The second model showed that the interaction effect of the patients' age and group did not have a significant effect on the blood glucose percentage change (*P*=.53). We then built a third model that included the same factors as the basic model and the interaction effect of the patients' education level and group. The third model showed that the interaction effect of the patients' education level and group did not have a significant effect on the blood glucose percentage change (*P*=.48). Finally, we built a fourth model with the same factors as the basic model and added the 2 interaction effects—one for the patient's age and group and another for the patient's education level and group. In the fourth model, we found that the interaction effects did not have a statistically significant effect on the blood glucose percentage change (*P*=.109 for the interaction effect of the patients' age and group; *P*=.94 for the interaction effect of the patients' education level and group). Therefore, we decided not to incorporate the age-group and education level-group interaction factors into the original model.

After conducting the above analyses, our original model was kept, and the final, included variables consisted of the patients' group, gender, diabetes duration, age, education level, the interaction effect of the nth month after registering an account, and the count of blood glucose measured in the nth month. Table 2 presents a summary of the new model.

Table 2. A summary of the new model.

Variable	Estimate	*t* test (*df*)	*P* value
Age	7.474×10^{-4}	2.652 (1.71×10^4)	.008
Education level[a]	2.927×10^{-2}	4.996 (1.72×10^4)	<.001
Patient groups[b]	-2.430×10^{-2}	–3.873 (1.81×10^4)	<.001
Gender[c]	2.753×10^{-3}	0.534 (1.74×10^4)	.59
Diabetes duration	1.010×10^{-3}	2.823 (1.71×10^4)	.005
Nth month	5.514×10^{-3}	2.212 (8.26×10^2)	.003
Count of blood glucose measured	1.403×10^{-4}	1.611 (2.37×10^4)	.11
Count of blood glucose measured in the nth month	-3.352×10^{-4}	–8.266 (1.97×10^4)	<.001

[a]The group of patients with an education level of junior high school or lower was set as the baseline.

[b]The only-data-uploader group was set as the baseline.

[c]The male patient cohort was set as the baseline.

We found that the app-engaged-user group had significantly greater decreases in blood glucose percentage change than the only-data-uploader group (β estimate=-2.430×10^{-2}; t=–3.873, df=1.81×10^4; P<.001 for the patients of the app-engaged-user group). In addition, for patients with shorter diabetes duration and those who are younger, the magnitudes of the drops in blood glucose percentage change were more profound (β estimate=1.010×10^{-3}; t=2.823, df=1.71×10^4; P=.005 for diabetes duration; β estimate=7.474×10^{-4}; t=2.652, df=1.71×10^4; P=.008 for the age of the patients; Figures 2-3). We also found that the frequency of SMBG enlarged the decreases in blood glucose along the interaction months (β estimate=-3.352×10^{-4}; t=–8.266, df=1.97×10^4; P<.001 for the nth month × the count of blood glucose in the nth month; Figure 4). Additionally, when the group of patients with an education level of junior high school or lower was set as the baseline, these patients had significantly greater decreases in blood glucose percentage change than those with an education level of senior high school or higher (β estimate=2.927×10^{-2}; t=4.996, df=1.72×10^4; P<.001 for patients with an education level of senior high school or higher; Figure 5). Lastly, the gender of the patients did not significantly affect the percentage change (β estimate=2.753×10^{-3}; t=0.534, df=1.74×10^4; P=.59 for female patients, with male patients as the baseline).

Figure 2. The relationship between blood glucose percentage changes and diabetes duration for each month as (A) a jittered scatter plot and (B) regression lines. In (A), the count of blood glucose measured in the nth month and the patients' age and educational level are fixed, and the overlaid regression lines are based on the estimated coefficients from the mixed model.

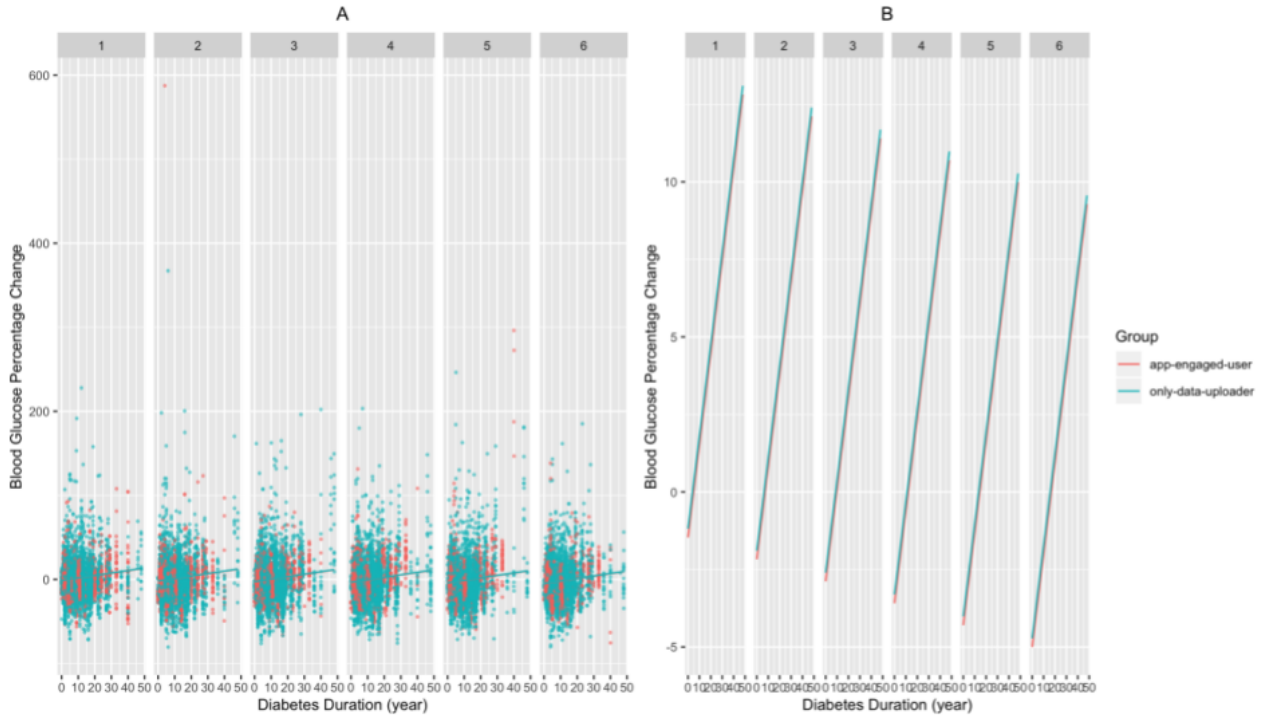


Figure 3. The relationship between blood glucose percentage changes and patient age for each month as (A) a jittered scatter plot and (B) regression lines. In (A), the count of blood glucose measured in the nth month and the patients' diabetes duration and educational level are fixed, and the overlaid regression lines are based on the estimated coefficients from the mixed model.

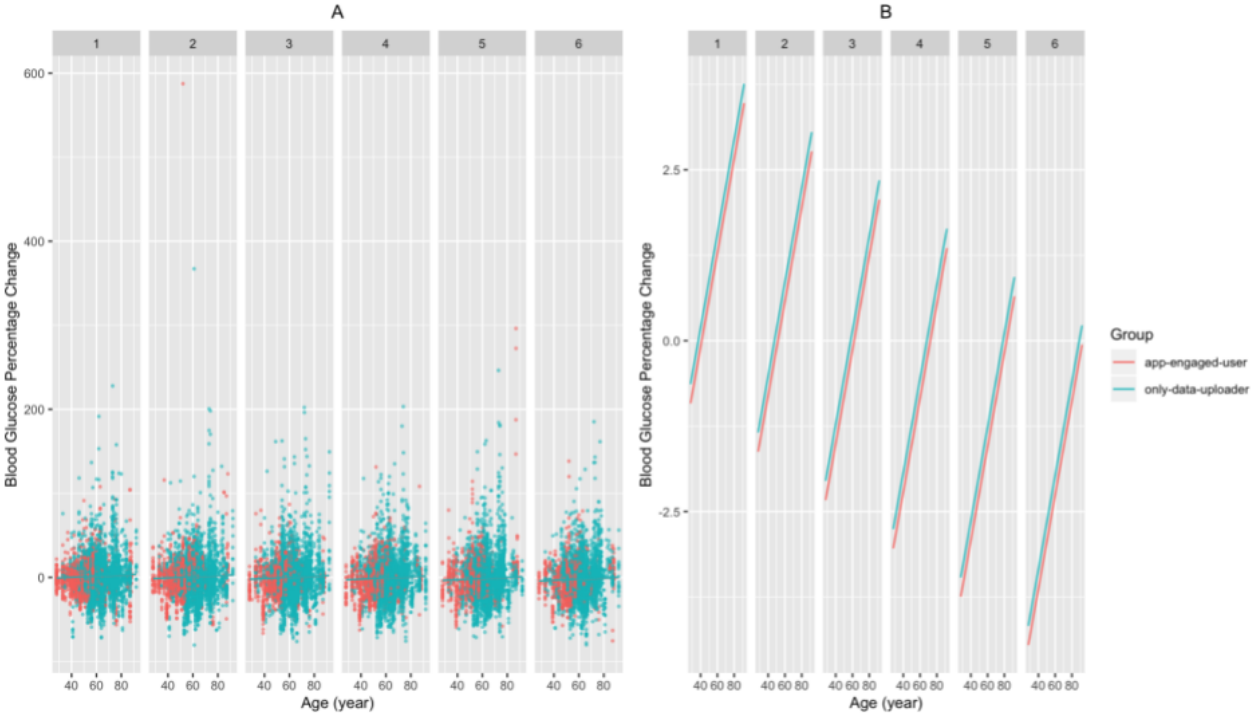


Figure 4. The relationship between blood glucose percentage changes and the count of blood glucose measured in each month as (A) a jittered scatter plot and (B) regression lines. In (A), the diabetes duration is fixed, and the overlaid regression lines are based on the estimated coefficients from the mixed model.

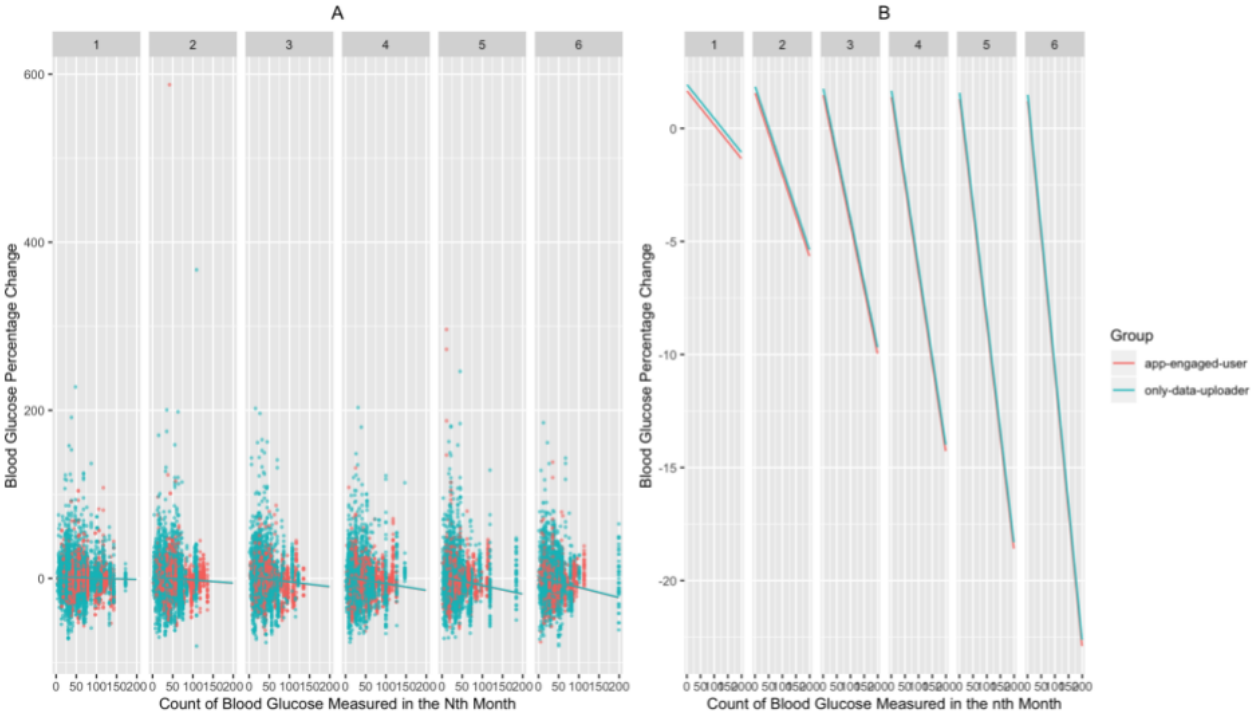
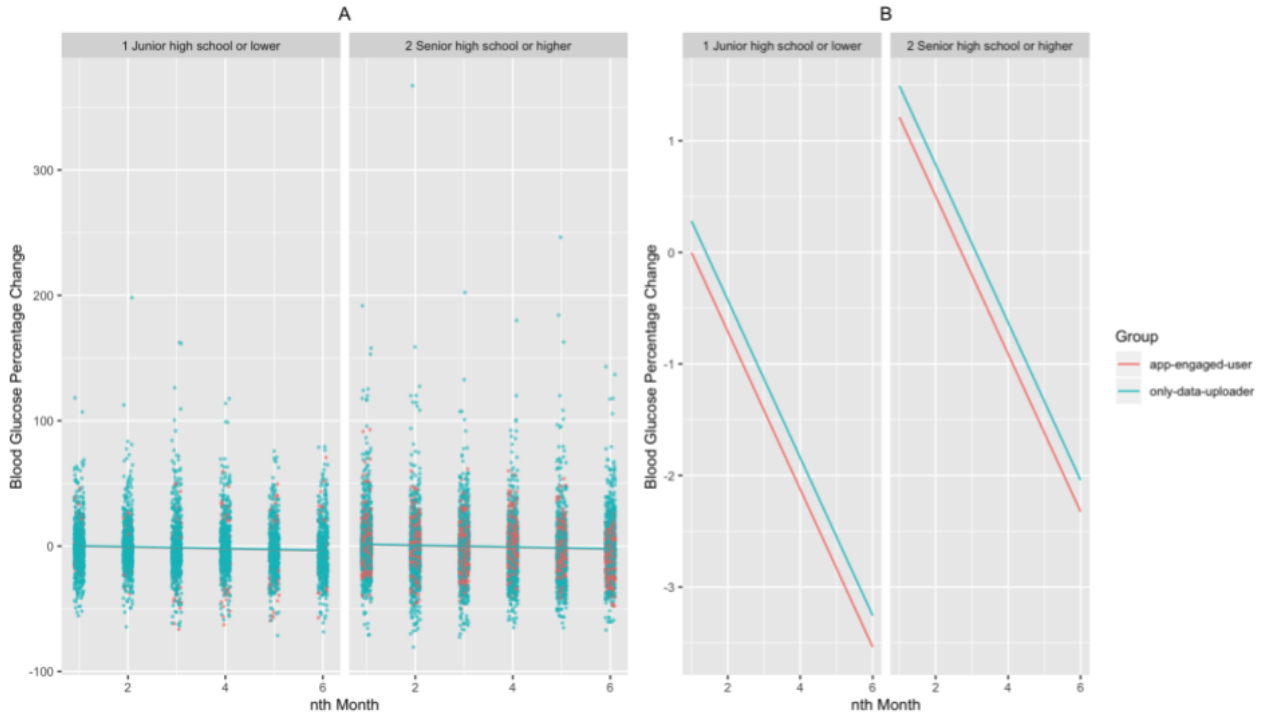


Figure 5. The relationship between blood glucose percentage changes and educational level for each month as (A) a jittered scatter plot and (B) regression lines. In (A), the count of blood glucose measured in the nth month and the patients' diabetes duration and age are fixed, and the overlaid regression lines are based on the estimated coefficients from the mixed model.



Discussion

Our study was based at a single clinic to minimize differences between the frequency of patient visits, level of health education, and quality of care received for the app-engaged-user and only-data-uploader groups. Our results showed that there were 6 significant factors—the patients' group (app-engaged-user or only-data-uploader), age, diabetes duration, education level, gender, and the count of blood glucose measured in the nth month—that were more strongly associated with changes in the patients' blood glucose. We found that patients who are app-engaged, younger, and less-educated and have shorter diabetes duration saw a steeper decrease in their blood glucose percentage change. We also found that the interaction between the nth month of recording SMBG and the SMBG count of that month affected blood glucose level significantly. Therefore, this interaction deserves more attention than total SMBG count. However, we found that the gender of the patients did not significantly affect the percentage change.

As previously mentioned, in many studies for patients with non–insulin-treated DMT2, the value of SMBG is inconsistent [8,11,22-24]. Some studies have demonstrated that SMBG was effective in controlling blood glucose [8,22,23,25], whereas other studies claimed that SMBG was not effective [12,26]. These inconsistencies are mainly due to differences in the trial designs, populations studied, and outcome indicators. However, in our findings, we used the count of blood glucose measured in the same month instead of the use of SMBG as a measurement. This is different from some previous studies. Diabetes patients test SMBG differently according to their current blood glucose status. Generally, when a patient's blood glucose becomes more stable, the count of SMBG will decrease. Therefore, it is more accurate to look at how the month of SMBG testing and count of SMBG each month affects blood glucose levels than total SMBG count.

Patients with longer diabetes duration may be affected by more diabetes symptoms [27-29], so their control of diabetes is usually worse than patients with a shorter diabetes duration. In addition, aging is associated with physiological changes that may lead to systemic alterations [30]. These systemic alterations may affect mental and physical functioning, increasing the chances of morbidity, multimorbidity, and mortality [30]. Older patients with diabetes may have macrovascular and microvascular complications and geriatric syndromes [28,31], so their control of diabetes is usually worse than younger patients with diabetes.

A common assumption is that patients with higher educational levels would have more knowledge about diseases and therapies, and thus, they would be able to better comply with therapies. However, previous studies have found that even highly educated patients may not sufficiently understand their conditions or truly believe in the benefits of therapy compliance, whereas patients with lower education levels may trust the doctor's advice more and exhibit better compliance [32,33]. This could explain our finding that the blood glucose percentage drop of the patients with an education level of junior high school or lower was greater than those with an education level of senior high school or higher. In addition, our study showed that the blood glucose percentage decrease of the patients who used the Health2Sync mobile app was more than those who did not use the app. For patients with the same education level, those using the Health2Sync mobile app had a greater decrease in blood glucose percentage than the patients who did not use it. Furthermore, regardless of the level of education, the patients who used the Health2Sync app experienced a larger drop in blood glucose

levels than those who do not use it. The Health2Sync app benefits users because it allows them to record their daily behaviors together with blood glucose readings, and the app has a bot that provides automated analyses, alerts, encouragements, and personalized educational content [16]. For the app-engaged-user group, 66% (69/104) of the patients recorded behavioral factors in addition to self-reported outcomes, most commonly entering diet, and 85% (88/104) viewed at least one educational content or interactive educational guide the app provided. Previous studies have shown that patient education and diet management are crucial for improving blood glucose [8,23,25,27]. There is growing evidence suggesting that gender affects the pathophysiology of many diseases, but in our study, gender did not significantly affect blood glucose percentage change [34-37]. Our study focused on a single clinic with limited samples; future studies should consider including a few more clinics to obtain more data samples for analyses. The other limitation is that we did not consider the differences in daily behaviors such as exercise and diet across the patients, and that these factors could have vital impacts on glycemic control. Future studies should also include these behaviors for analyses.

Diabetes is approaching epidemic proportions globally, and it places an enormous burden upon both the patients and countries' health systems. It is especially difficult for low- and middle-income countries, due to insufficient equipment and clinics, to cope with the rise in diabetes and other chronic diseases [25,38]. The Health2Sync app can enhance the care for patients with diabetes and solve resource-limited problems.

Additionally, our study showed positive results at a single diabetes management clinic using real-world data without prior RCT settings. RCTs are generally considered by regulators to be the gold standard for establishing the causal relationship between medication and patient outcomes, but it is incapable of reflecting real clinical practice in which heterogeneous scenarios exist [39-41]. As digital interventions are to be applied to all patients, we believe that our study with real-world data is more convincing in demonstrating efficacy.

In conclusion, through the retrospective analyses, we showed that the Health2Sync app and SMBG contribute to the improvement and control of blood glucose. Further studies are needed to reveal whether different clinical care methods have an impact on diabetes treatment.

Conflicts of Interest

The Health2Sync mobile app and web-based Patient Management Platform are products of H2 Inc. YTC, YZT, and KL are full-time employees at H2 Inc, and KL supervises YZT and YTC. HYC received a consulting fee to assist with the analyses but otherwise declared no conflict of interest. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
List of blood glucose meters used by patients.
[DOCX File , 13 KB - mhealth_v10i6e31764_app1.docx]

References

1. Harding J, Shaw J, Peeters A, Cartensen B, Magliano D. Cancer risk among people with type 1 and type 2 diabetes: disentangling true associations, detection bias, and reverse causation. Diabetes Care 2015 Feb;38(2):264-270. [doi: 10.2337/dc14-1996] [Medline: 25488912]
2. Cho NH, Shaw J, Karuranga S, Huang Y, da Rocha Fernandes J, Ohlrogge A, et al. IDF diabetes atlas: global estimates of diabetes prevalence for 2017 and projections for 2045. Diabetes Res Clin Pract 2018 Apr;138:271-281. [doi: 10.1016/j.diabres.2018.02.023] [Medline: 29496507]
3. Baena-Díez D, Jose M, Peñafiel J, Subirana I, Ramos R, Elosua R, FRESCO Investigators. Risk of cause-specific death in individuals with diabetes: a competing risks analysis. Diabetes Care 2016 Nov;39(11):1987-1995. [doi: 10.2337/dc16-0614] [Medline: 27493134]
4. Goldstein DE, Little RR, Lorenz RA, Malone JI, Nathan DM, Peterson CM, et al. Tests of glycemia in diabetes. Diabetes Care 2004 Jul 01;27(7):1761-1773 [FREE Full text] [doi: 10.2337/diacare.27.7.1761] [Medline: 15220264]
5. Franciosi M, Pellegrini F, De Berardis G, Belfiglio M, Di Nardo B, Greenfield S, QuED Study Group--Quality of CareOutcomes in Type 2 Diabetes. Self-monitoring of blood glucose in non-insulin-treated diabetic patients: a longitudinal evaluation of its impact on metabolic control. Diabet Med 2005 Jul;22(7):900-906. [doi: 10.1111/j.1464-5491.2005.01546.x] [Medline: 15975106]
6. Wada E, Onoue T, Kobayashi T, Handa T, Hayase A, Ito M, et al. Flash glucose monitoring helps achieve better glycemic control than conventional self-monitoring of blood glucose in non-insulin-treated type 2 diabetes: a randomized controlled trial. BMJ Open Diabetes Res Care 2020 Jun;8(1):e001115 [FREE Full text] [doi: 10.1136/bmjdrc-2019-001115] [Medline: 32518063]
7. Chen C, Hung L, Chen Y, Yeh MC. Perspectives of patients with non-insulin-treated type 2 diabetes on self-monitoring of blood glucose: a qualitative study. J Clin Nurs 2018 Apr 26;27(7-8):1673-1683. [doi: 10.1111/jocn.14227] [Medline: 29266453]

8. Parkin CG, Buskirk A, Hinnen DA, Axel-Schweitzer M. Results that matter: structured vs. unstructured self-monitoring of blood glucose in type 2 diabetes. Diabetes Res Clin Pract 2012 Jul;97(1):6-15. [doi: 10.1016/j.diabres.2012.03.002] [Medline: 22464874]
9. Mannucci E, Antenore A, Giorgino F, Scavini M. Effects of structured versus unstructured self-monitoring of blood glucose on glucose control in patients with non-insulin-treated type 2 diabetes: a meta-analysis of randomized controlled trials. J Diabetes Sci Technol 2018 Jan 01;12(1):183-189 [FREE Full text] [doi: 10.1177/1932296817719290] [Medline: 28697625]
10. Xu Y, Tan D, Lee J. Evaluating the impact of self-monitoring of blood glucose frequencies on glucose control in patients with type 2 diabetes who do not use insulin: a systematic review and meta-analysis. Int J Clin Pract 2019 Jul;73(7):e13357 [FREE Full text] [doi: 10.1111/ijcp.13357] [Medline: 31033116]
11. Williams DM, Parsons SN, Dunseath GJ, Stephens JW, Luzio SD, Owens DR. The impact of structured self-monitoring of blood glucose on glycaemic variability in non-insulin treated type 2 diabetes: the SMBG study, a 12-month randomised controlled trial. Diabetes Metab Syndr 2020 Mar;14(2):101-106. [doi: 10.1016/j.dsx.2020.01.006] [Medline: 31995784]
12. Young LA, Buse JB, Weaver MA, Vu MB, Mitchell CM, Blakeney T, Monitor Trial Group. Glucose self-monitoring in non-insulin-treated patients with type 2 diabetes in primary care settings: a randomized trial. JAMA Intern Med 2017 Jul 01;177(7):920-929 [FREE Full text] [doi: 10.1001/jamainternmed.2017.1233] [Medline: 28600913]
13. Zhu H, Zhu Y, Leung S. Is self-monitoring of blood glucose effective in improving glycaemic control in type 2 diabetes without insulin treatment: a meta-analysis of randomised controlled trials. BMJ Open 2016 Sep 02;6(9):e010524 [FREE Full text] [doi: 10.1136/bmjopen-2015-010524] [Medline: 27591016]
14. Davis W, Bruce D, Davis T. Is self-monitoring of blood glucose appropriate for all type 2 diabetic patients? the Fremantle Diabetes Study. Diabetes Care 2006 Aug;29(8):1764-1770. [doi: 10.2337/dc06-0268] [Medline: 16873777]
15. Karter A, Parker M, Moffet H, Spence MM, Chan J, Ettner SL, et al. Longitudinal study of new and prevalent use of self-monitoring of blood glucose. Diabetes Care 2006 Aug;29(8):1757-1763 [FREE Full text] [doi: 10.2337/dc06-2073] [Medline: 16873776]
16. Tu YZ, Chang YT, Chiou HY, Lai K. The effects of continuous usage of a diabetes management app on glycemic control in real-world clinical practice: retrospective analysis. J Med Internet Res 2021 Jul 15;23(7):e23227 [FREE Full text] [doi: 10.2196/23227] [Medline: 34264192]
17. Vegesna A, Tran M, Angelaccio M, Arcona S. Remote patient monitoring via non-invasive digital technologies: a systematic review. Telemed J E Health 2017 Jan;23(1):3-17 [FREE Full text] [doi: 10.1089/tmj.2016.0051] [Medline: 27116181]
18. ISO 15197:2013(en) In vitro diagnostic test systems — requirements for blood-glucose monitoring systems for self-testing in managing diabetes mellitus. International Organization for Standardization. 2013 May. URL: https://www.iso.org/obp/ui/#iso:std:iso:15197:ed-2:v1:en [accessed 2022-05-11]
19. R Core Team. R: a language and environment for statistical computing. The R Project for Statistical Computing. 2019. URL: https://www.R-project.org/ [accessed 2022-05-10]
20. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Soft 2015 Oct 07;67(1):1-48. [doi: 10.18637/jss.v067.i01]
21. Verbyla A. A note on model selection using information criteria for general linear models estimated using REML. Aust N Z J Stat 2019 Mar 09;61(1):39-50 [FREE Full text] [doi: 10.1111/anzs.12254]
22. Schütt M, Kern W, Krause U, Busch P, Dapp A, Grziwotz R, et al. Is the frequency of self-monitoring of blood glucose related to long-term metabolic control? multicenter analysis including 24,500 patients from 191 centers in Germany and Austria. Exp Clin Endocrinol Diabetes 2006 Jul 16;114(7):384-388. [doi: 10.1055/s-2006-924152] [Medline: 16915542]
23. Barnett AH, Krentz AJ, Strojek K, Sieradzki J, Azizi F, Embong M, et al. The efficacy of self-monitoring of blood glucose in the management of patients with type 2 diabetes treated with a gliclazide modified release-based regimen. a multicentre, randomized, parallel-group, 6-month evaluation (DINAMIC 1 study). Diabetes Obes Metab 2008 Dec 20;10(12):1239-1247. [doi: 10.1111/j.1463-1326.2008.00894.x] [Medline: 18494813]
24. Polonsky WH, Fisher L, Schikman CH, Hinnen DA, Parkin CG, Jelsovsky Z, et al. Structured self-monitoring of blood glucose significantly reduces A1C levels in poorly controlled, noninsulin-treated type 2 diabetes: results from the Structured Testing Program study. Diabetes Care 2011 Feb 26;34(2):262-267 [FREE Full text] [doi: 10.2337/dc10-1732] [Medline: 21270183]
25. Pillay S, Aldous C. Effects of self-monitoring of blood glucose on diabetes control in a resource-limited diabetic clinic. J Endocrinol Metab Diabetes South Afr 2016 Jun 27;21(2):20-25. [doi: 10.1080/16089677.2016.1198112]
26. Sodipo OO, Adedokun A, Olusola AA. Effect of self-monitoring of blood glucose on glycaemic outcome among type 2 diabetic patients. S Afr Fam Pract 2017 Dec 05;59(2):35. [doi: 10.4102/safp.v59i6.4643]
27. Ko S, Park S, Cho J, Ko S, Shin K, Lee S, et al. Influence of the duration of diabetes on the outcome of a diabetes self-management education program. Diabetes Metab J 2012 Jun;36(3):222-229 [FREE Full text] [doi: 10.4093/dmj.2012.36.3.222] [Medline: 22737662]
28. Ito H, Omoto T, Abe M, Matsumoto S, Shinozaki M, Nishio S, et al. Relationships between the duration of illness and the current status of diabetes in elderly patients with type 2 diabetes mellitus. Geriatr Gerontol Int 2017 Jan 04;17(1):24-30. [doi: 10.1111/ggi.12654] [Medline: 26634299]

29. Kasteleyn MJ, de Vries L, van Puffelen AL, Schellevis FG, Rijken M, Vos RC, Diacourse study group. Diabetes-related distress over the course of illness: results from the Diacourse study. Diabet Med 2015 Dec 28;32(12):1617-1624. [doi: 10.1111/dme.12743] [Medline: 25763843]
30. World report on ageing and health. World Health Organization. 2015. URL: https://apps.who.int/iris/bitstream/handle/10665/186463/9789240694811_eng.pdf?sequence=1&isAllowed=y [accessed 2022-05-10]
31. Vasconcelos C, Almeida A, Cabral M, Ramos E, Mendes R. The impact of a community-based food education program on nutrition-related knowledge in middle-aged and older patients with type 2 diabetes: results of a pilot randomized controlled trial. Int J Environ Res Public Health 2019 Jul 06;16(13):2403 [FREE Full text] [doi: 10.3390/ijerph16132403] [Medline: 31284568]
32. Jin J, Sklar GE, Min Sen Oh V, Chuen Li S. Factors affecting therapeutic compliance: A review from the patient's perspective. Ther Clin Risk Manag 2008 Feb;4(1):269-286 [FREE Full text] [doi: 10.2147/tcrm.s1458] [Medline: 18728716]
33. Senior V, Marteau TM, Weinman J, Genetic Risk Assessment for FH Trial (GRAFT) Study Group. Self-reported adherence to cholesterol-lowering medication in patients with familial hypercholesterolaemia: the role of illness perceptions. Cardiovasc Drugs Ther 2004 Nov;18(6):475-481. [doi: 10.1007/s10557-004-6225-z] [Medline: 15770435]
34. Mauvais-Jarvis F. Gender differences in glucose homeostasis and diabetes. Physiol Behav 2018 Apr 01;187:20-23 [FREE Full text] [doi: 10.1016/j.physbeh.2017.08.016] [Medline: 28843891]
35. van Genugten RE, Utzschneider KM, Tong J, Gerchman F, Zraika S, Udayasankar J, American Diabetes Association GENNID Study Group. Effects of sex and hormone replacement therapy use on the prevalence of isolated impaired fasting glucose and isolated impaired glucose tolerance in subjects with a family history of type 2 diabetes. Diabetes 2006 Dec;55(12):3529-3535 [FREE Full text] [doi: 10.2337/db06-0577] [Medline: 17130501]
36. Mauvais-Jarvis F, Bairey Merz N, Barnes PJ, Brinton RD, Carrero J, DeMeo DL, et al. Sex and gender: modifiers of health, disease, and medicine. Lancet 2020 Aug 22;396(10250):565-582 [FREE Full text] [doi: 10.1016/S0140-6736(20)31561-0] [Medline: 32828189]
37. Tonolo G. Sex-gender awareness in diabetes. Diabetology 2021 Jun 21;2(2):117-122. [doi: 10.3390/diabetology2020010]
38. Duffy S, Svenson J, Chavez A, Kelly M, Wise P. Empowering community health workers with mobile technology to treat diabetes. Ann Fam Med 2019 Mar 11;17(2):176-176 [FREE Full text] [doi: 10.1370/afm.2361] [Medline: 30858263]
39. Franklin J, Schneeweiss S. When and how can real world data analyses substitute for randomized controlled trials? Clin Pharmacol Ther 2017 Dec;102(6):924-933 [FREE Full text] [doi: 10.1002/cpt.857] [Medline: 28836267]
40. Makady A, de Boer A, Hillege H, Klungel O, Goettsch W, (on behalf of GetReal Work Package 1). What is real-world data? a review of definitions based on literature and stakeholder interviews. Value Health 2017 Jul;20(7):858-865 [FREE Full text] [doi: 10.1016/j.jval.2017.03.008] [Medline: 28712614]
41. Mahajan R. Real world data: additional source for making clinical decisions. Int J Appl Basic Med Res 2015;5(2):82 [FREE Full text] [doi: 10.4103/2229-516X.157148] [Medline: 26097811]

Abbreviations

DMT2: diabetes mellitus type 2
FBG: fasting blood glucose
HCP: health care professional
RCT: randomized controlled trial
SMBG: self-monitoring of blood glucose

Edited by L Buis, A Mavragani; submitted 03.07.21; peer-reviewed by M Khokhar, Anonymous; comments to author 27.08.21; revised version received 23.12.21; accepted 21.04.22; published 15.06.22.

Please cite as:
Chang YT, Tu YZ, Chiou HY, Lai K, Yu NC
Real-world Benefits of Diabetes Management App Use and Self-monitoring of Blood Glucose on Glycemic Control: Retrospective Analyses
JMIR Mhealth Uhealth 2022;10(6):e31764
URL: https://mhealth.jmir.org/2022/6/e31764
doi:10.2196/31764
PMID:35704371

Original Paper

Evaluating the Efficacy of the Drinks:Ration Mobile App to Reduce Alcohol Consumption in a Help-Seeking Military Veteran Population: Randomized Controlled Trial

Daniel Leightley[1*], BSc, MSc, PhD; Charlotte Williamson[1*], BSc, MSc; Roberto J Rona[1], PhD; Ewan Carr[2], PhD; James Shearer[3], PhD; Jordan P Davis[4], PhD; Amos Simms[5,6], BSc, MSc; Nicola T Fear[1,5], BSc, MSc, DPhil; Laura Goodwin[7*], BSc, MSc, PhD; Dominic Murphy[1,8*], MA, PhD, DCM

[1]Institute of Psychiatry, Psychology & Neuroscience, King's Centre for Military Health Research, King's College London, London, United Kingdom
[2]Department of Biostatistics and Health Informatics, King's College London, London, United Kingdom
[3]King's Health Economics, King's College London, London, United Kingdom
[4]Suzanne Dworak-Peck School of Social Work, University of Southern California, Los Angeles, CA, United States
[5]Academic Department of Military Mental Health, King's College London, London, United Kingdom
[6]British Army, London, United Kingdom
[7]Division of Health Research, Lancaster University, Lancaster, United Kingdom
[8]Combat Stress, Leatherhead, United Kingdom
[*]these authors contributed equally

Corresponding Author:
Daniel Leightley, BSc, MSc, PhD
Institute of Psychiatry, Psychology & Neuroscience
King's Centre for Military Health Research
King's College London
10 Cutcombe Road
London, SE5 9RJ
United Kingdom
Phone: 44 78485351
Email: daniel.leightley@kcl.ac.uk

Abstract

Background: Alcohol misuse is higher in the UK armed forces (AF) than in the general population. Research demonstrates that alcohol misuse persists after an individual leaves service, and this is notably the case for those who are seeking help for a mental health difficulty. Despite this, there is no work on testing a mobile alcohol reduction intervention that is personalized to support the UK AF.

Objective: To address this gap, we investigated the efficacy of a 28-day brief alcohol intervention delivered via a mobile app in reducing weekly self-reported alcohol consumption among UK veterans seeking help for mental health difficulties.

Methods: We performed a 2-arm participant-blinded randomized controlled trial (RCT). We compared a mobile app that included interactive features designed to enhance participants' motivation and personalized messaging (intervention arm) with a version that provided government guidance on alcohol consumption only (control arm). Adults were eligible if they had served in the UK AF, were currently receiving or had received clinical support for mental health symptoms, and consumed 14 units (approximately 112 g of ethanol) or more of alcohol per week. Participants received the intervention or the control mobile app (1:1 ratio). The primary outcome was a change in self-reported weekly alcohol consumption between baseline and day 84 assessed using the validated Timeline Follow Back for Alcohol Consumption (TLFB) (prior 7 days), with a secondary outcome exploring self-reported change in the Alcohol Use Disorder Identification Test (AUDIT) score.

Results: Between October 2020 and April 2021, 2708 individuals were invited to take part, of which 2531 (93.5%) did not respond, 54 (2%) were ineligible, and 123 (4.5%) responded and were randomly allocated (62, 50.4%, intervention; 61, 49.6%, control). At day 84, 41 (66.1%) participants in the intervention arm and 37 (60.7%) in the control arm completed the primary outcome assessment. Between baseline and day 84, weekly alcohol consumption reduced by –10.5 (95% CI –19.5 to –1.5) units in the control arm and –28.2 (95% CI –36.9 to –19.5) units in the intervention arm (P=.003, Cohen d=0.35). We also found a

significant reduction in the AUDIT score of –3.9 (95% CI –6.2 to –1.6) in the intervention arm (Cohen *d*=0.48). Our primary and secondary effects did not persist over the longer term (day 168). Two adverse events were detected during the trial.

Conclusions: This study examined the efficacy of a fully automated 28-day brief alcohol intervention delivered via a mobile app in a help-seeking sample of UK veterans with hazardous alcohol consumption. We found that participants receiving *Drinks*:Ration reduced their alcohol consumption more than participants receiving guidance only (at day 84). In the short term, we found *Drinks*:Ration is efficacious in reducing alcohol consumption in help-seeking veterans.

Trial Registration: ClinicalTrials.gov NCT04494594; https://tinyurl.com/34em6n9f

International Registered Report Identifier (IRRID): RR2-10.2196/19720

(JMIR Mhealth Uhealth 2022;10(6):e38991) doi:10.2196/38991

KEYWORDS

military; veteran; digital health; alcohol misuse; smartphone; mobile health; mHealth; alcohol intervention; digital intervention; mental health; smartphone application; health intervention; alcohol consumption

Introduction

Evidence has shown alcohol misuse is substantially higher in the United Kingdom's armed forces (AF) than the UK general population [1,2]. Research has demonstrated that alcohol misuse persists after leaving service [3], and 43% of veterans seeking treatment for a mental health difficulty report misusing alcohol [4]. In the United Kingdom, a veteran is defined as an individual serving at least 1 day of paid employment in the AF. Alcohol misuse often co-occurs with posttraumatic stress disorder (PTSD), anxiety, or depression and frequently used as a coping mechanism [5].

Research has shown that help-seeking veterans (those seeking support in a clinical setting) misusing alcohol attend fewer mental health appointments and are more likely to have a negative perception of mental health treatment [6]. This is, in part, due to those misusing alcohol being denied access to mental health treatment services until they have reduced the hazardous drinking. Thus, interventions that target drinking behavior need to be developed as this may enhance engagement with mental health services and improve mental health outcomes and quality of life.

The past 5 years have seen a growing treatment gap, with patients waiting longer for mental health referrals and treatment in the United Kingdom. This has further been exacerbated by the COVID-19 pandemic. To address these issues, modes of intervention delivery have shifted from in-person to web-based to mobile-based delivery [7]. Mobile interventions for alcohol misuse in the United Kingdom, such as Drink Less [8] and Drinkaware [9], have several advantages over web-based delivery, including (1) more holistic delivery of behavior changes, (2) the use of mobile sensors and wearables to inform decision-making, (3) avoiding the stigma associated with receiving help in person, and (4) convenience since they can be used when the individual prefers (discretely or openly). Mobile interventions also offer a more cost-efficient way to deliver behavior change techniques (BCTs, the specific and active component of an intervention designed to change behavior [10]) for reducing alcohol use.

Existing alcohol apps targeted at the general public include self-monitoring apps (eg, Drink Less [8], Drinkaware [9], One You Drinks Tracker [11]), where users are encouraged to regularly record and monitor (via visualizations) their alcohol consumption. Self-monitoring has been found to be associated with improved outcomes and is an effective BCT for reducing alcohol use. A recent review of personalized digital interventions found reductions in hazardous and harmful alcohol consumption to be associated with behavior substitution, problem solving, and providing a credible source of information [12]. Another review also identified the role that personalized notifications play in promoting positive changes in behavior [13]. However, current mobile interventions focused on the general population do not target aspects experienced by the AF community, such as individual beliefs, prevailing social context, comorbid mental health problems, military service experience, and perceptions of consumption [14]. Further, existing apps do not cater for the excessive amounts of alcohol consumed by UK AF personnel.

To date, there is no published work that seeks to test a brief automated mobile intervention alcohol reduction app that is personalized to support UK AF, considering their military experiences. To address this, we developed the *Drinks*:Ration app (previously called Information about Drinking for Ex-Serving personnel [InDEx]; see [15-18]) to support UK AF veterans to reduce the amount they drink.

We conducted a randomized controlled trial (RCT) to assess the efficacy of a 28-day alcohol intervention delivered via *Drinks*:Ration in reducing self-reported weekly alcohol consumption by day 84 follow-up among veterans who drink at hazardous or harmful levels (ie, drinking at a level likely to cause harm) and are currently receiving, or have previously received, support for mental health symptoms in a clinical setting.

Methods

Study Design and Hypotheses

This was a 2-arm participant-blinded (single-blinded) RCT (1:1) to compare a mobile app that provided government guidance on alcohol consumption only (control arm) with the mobile app *Drinks*:Ration, a personalized app based on BCT principles. We hypothesized that the intervention arm would be efficacious in reducing alcohol consumption when compared to the control arm.

Both the control and intervention arms were delivered via the *Drinks*:Ration app. Participants in the control arm were given access only to the alcohol consumption guidance and a unit calculator based on guidance issued by the UK Chief Medical Officer. Those in the intervention arm were given access to the full version of the app, which included theoretically driven components and personalized messaging. This included individualized normative guidance alongside features designed to enhance participants' motivation through interactive guidance focused on self-efficacy to help modify alcohol consumption. Participants in both arms were asked to use the app for a minimum of 28 days. After this, they could continue to use the app, but they did not receive personalized messaging. This was undertaken to assess the long-term effectiveness of the app.

This study was designed such that the control arm structurally resembled the intervention arm but excluded active intervention techniques, such as a drinks diary, drinks in pixels, and drinking zones (based on the Global Positioning System [GPS] location). This approach increased uniformity across the arms (eg, ensuring both arms received a digital intervention) and maintained treatment allocation concealment.

Data were collected at baseline and follow-up assessments at 84 and 168 days postbaseline. Additional questionnaires were collected on days 7, 14, and 21. This information was used to personalize the *Drinks*:Ration app for the intervention arm only. Please refer to the published trial protocol for further details [19].

Ethical Considerations

This trial was approved by the local ethics committee of King's College London (HR-19/20-17438).

Procedure and Participants

Participants were recruited between October 2020 and April 2021 in succession via a clinical group, an existing research cohort [2], and social media [20]. The clinical group was derived from Combat Stress, a third-sector charitable organization that provides mental health services, including substance misuse management, to UK veterans. The research cohort was the King's Centre for Military Health Research health and well-being longitudinal cohort study [2], where a sample of self-reported help seekers were identified and extracted. Social media platforms, such as Facebook and Twitter, were also used to promote the RCT via free and paid promotional advertisements with a link enabling potential users to express an interest in taking part (for further information, please see [20]).

Potential participants were invited to take part via email with an explanation of the study, a link to the participant information sheet, and instructions on how to download *Drinks*:Ration using a unique quick response (QR) code. Once participants had downloaded the app, they were asked to report alcohol consumption using the validated Timeline Follow Back for Alcohol Consumption (TLFB) [21] for the prior 7 days and confirm their military serving status. Those meeting the study eligibility criteria were allowed to proceed and complete the baseline questionnaire.

Eligibility was assessed at baseline. To be included in this RCT, participants needed to download the *Drinks*:Ration app onto an iOS or Android device; be aged 18 years or older; currently reside in the United Kingdom; consume at least 14 UK units (approximately 112 g of ethanol) of alcohol or more per week at baseline (hazardous or harmful levels of alcohol consumption); confirm that they currently receive, or had received, support for mental health symptoms in a clinical setting; provide a mobile phone number; and be a veteran of the UK AF.

Sample Size

A power calculation was performed based on Alcohol Use Disorder Identification Test (AUDIT) data previously reported from Combat Stress [5]. To detect a difference in alcohol consumption of 4 UK units (approximately 40 g of alcohol per week) between the control and intervention arms at day 84, with a 2-sided 5% significance level and a power of 80%, we needed a sample of 37 participants per arm with complete primary outcome assessments. We selected 4 UK units based on reductions observed in similar studies [8,22,23] and reductions observed in the feasibility trial of *Drinks*:Ration, which found a 7-unit decrease at week 4. To allow for attrition of 40%, we aimed to recruit a total of 124 participants (62, 50%, per arm). To account for an expected response rate of 30%, we estimated that we would need to invite 620 veterans to participate in the study.

Randomization and Masking

Randomization was carried out automatically as part of the *Drinks*:Ration platform. When participants registered for the app, they were assigned a unique identifier and asked to provide their gender. They were then randomly allocated [1:1] to receive either the control or the intervention arm. Stratification was used to ensure equal gender distribution across arms. This is because those who identify as female only represent approximately 10% of the UK AF.

The randomization procedure was based on a list of random numbers computer-generated by the *Drinks*:Ration platform. All members of the research team were blind to participant treatment allocations except for authors DL and CW. CW conducted participant management, and DL led the development of the *Drinks*:Ration app, had access to raw study data, and conducted the primary analyses. Except for automated weekly backups, access to the data was disabled.

Participants were not informed of their treatment allocation. However, they may have been able to deduce their allocation condition based on app content. As the intervention was automated and delivered via an app, there was no contact between researchers and participants during the intervention unless a risk to health (adverse event) had been detected or if technical problems arose [19]. An adverse event in this study was defined as participants reporting (via the drinks diary or during contact with the research team) that they had consumed more than 25 UK units of alcohol within a 24-hour period. Once detected, a clinician would contact the participant over the telephone to check their health and provide signposting to other

services (which are listed in the app). The clinician would not disclosure the treatment arm allocation.

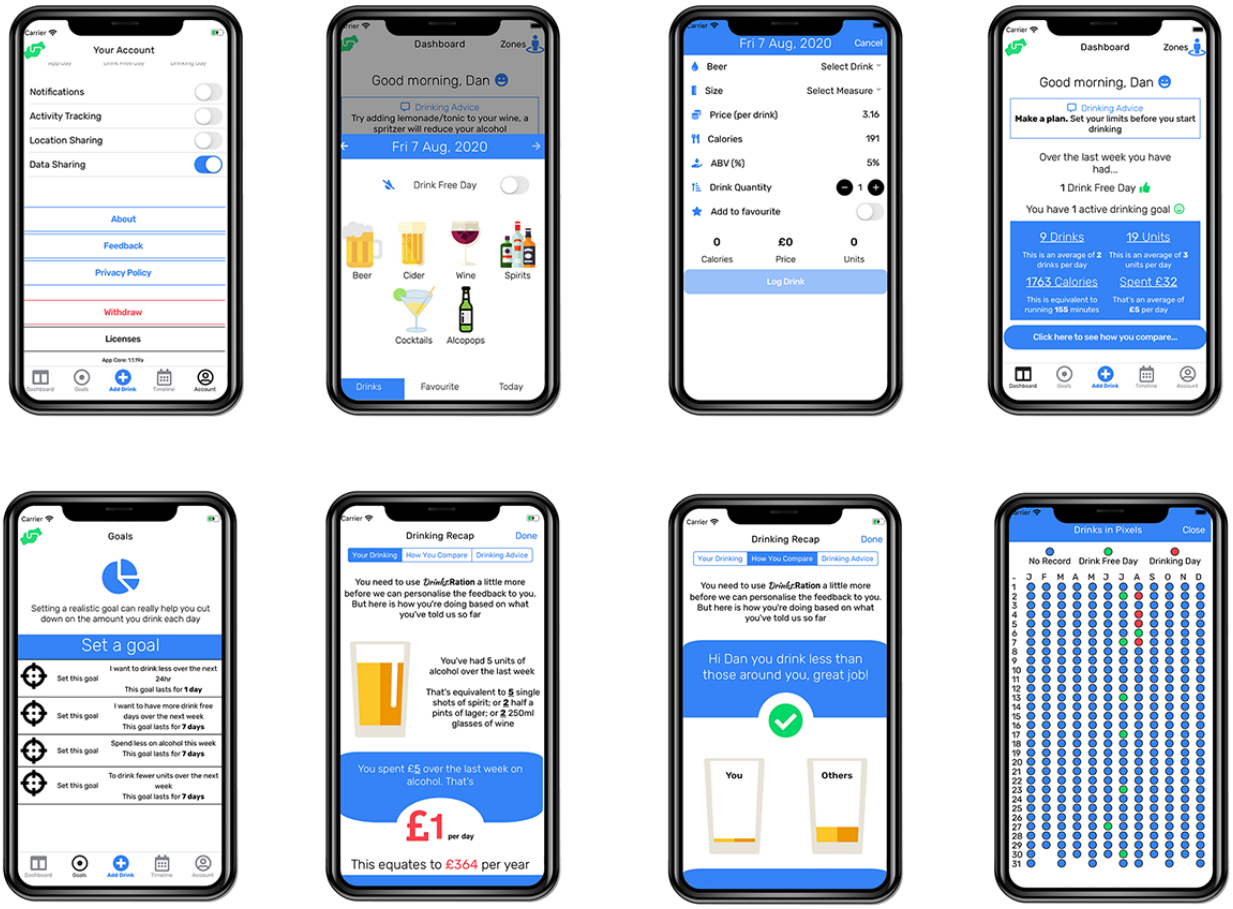
Intervention

Drinks:Ration (formerly called InDEx [15-17]) was developed by the King's Centre for Military Health Research (at King's College London) and the University of Liverpool following the Medical Research Council Complex Intervention Guidelines using co-design methodology in collaboration with end users. The app was designed to support veterans drinking at hazardous or harmful levels by providing bespoke advice and support over a minimum of 28 days. The app was developed without any organizational branding to promote use. The iterative development process, theoretical framework, and feasibility trial are reported elsewhere [15-17,19]. Briefly, *Drinks*:Ration was developed and tested with 5 core modules (see Figure 1):

- Account management: Participants can modify personal information (eg, first name and mobile number) and app parameters (eg, automatic logout, clear local storage, data sharing permission, and leaving the study).
- Questionnaires and individualized normative guidance: This captures the participant's response to a set of questions and aggregates responses to produce an individualized infographic representing the participant's alcohol consumption in comparison to the general population, the AF community, and other participants of the *Drinks*:Ration app.
- Self-monitoring and guidance: This records alcohol consumption by participants and provides a range of visualizations to allow consumption monitoring. Further, participants can customize the visualizations with metrics they find relevant (eg, calories, cost, or exercise required).
- Goals (setting and review): Participants can set goals based on the implementation intentions (if and then) methodology; visualizations provide guidance on progress toward achieving specified goals.
- Personalized messaging: Participants are sent tailored messages via push notifications or short message service (SMS) text messages that provide prompts to use the drinks diary, suggest alternative behaviors, and provide guidance on goals (see Multimedia Appendix 1 for example messaging).

Participants in the intervention arm completed additional questionnaires on their mood and general mental health each week. These responses were used to personalize app content, push notifications, and SMS text messages.

Figure 1. Example screenshots extracted from the *Drinks*:Ration app.



Measures

A summary of measures and data collection schedule is provided in Table 1 (and see [19]). Measures were the same for the intervention and control arms.

Table 1. Summary of measures and data collection timepoints. Days 7, 14, and 21 measures were used to personalize the *Drinks*:Ration app and apply to the intervention arm only.

Measure	Day 0 (baseline)	Day 7	Day 14	Day 21	Day 28	Day 84 (primary endpoint)	Day 168
Informed consent	I[a]/C[b]	N/A[c]	N/A	N/A	N/A	N/A	N/A
Sociodemographics	I/C	N/A	N/A	N/A	N/A	N/A	N/A
Depression (2-item Patient Health Questionnaire [PHQ2]) [24]	I/C	I/C	I/C	I/C	I/C	I/C	I/C
Anxiety (2-item Generalized Anxiety Disorder [GAD2]) [25]	I/C	I/C	I/C	I/C	I/C	I/C	I/C
International Trauma Questionnaire (ITQ) for PTSD[d] [26]	I/C	N/A	N/A	N/A	I/C	I/C	I/C
AUDIT[e] [27]	I/C	N/A	N/A	N/A	I/C	I/C	I/C
7-day TLFB[f] [21]	I/C	N/A	N/A	N/A	I/C	I/C	I/C
Usability evaluation (MAUQ[g]) [28]	N/A	N/A	N/A	N/A	I/C	N/A	N/A

[a]I: intervention arm.

[b]C: control arm.

[c]N/A: not applicable.

[d]PTSD: posttraumatic stress disorder.

[e]AUDIT: Alcohol Use Disorder Identification Test (10 items).

[f]TLFB: Timeline Follow Back for Alcohol Consumption.

[g]MAUQ: mobile health (mHealth) App Usability Questionnaire.

Baseline Measures

Upon successful registration, participants completed a baseline questionnaire to assess physical and mental health, health status, resource utilization, and sociodemographics.

Outcome Measures

The primary outcome was a change between baseline and day 84 follow-up in self-reported alcohol consumption, as measured by the 7-day TLFB. Participants were asked to report how many drinks they consumed over the past 7 days, as well as the type of drink consumed each day. Using standard unit measurements (see Multimedia Appendix 2 for an outline), the weekly alcohol consumption for baseline, day 28, day 84, and day 168 was determined by summing the number of units assigned to each drink.

The secondary outcome measure assessed changes in the AUDIT score from baseline to day 84 follow-up. The day 84 follow-up timepoint was selected to assess the short- to medium-term benefits of the intervention, although outcomes were also examined at day 168 to assess longer-term benefits. Changes in quality of life (eg, physical health, psychological health, social relationships, and environment) and cost-effectiveness will be analyzed in future papers.

At each follow-up, participants were first asked to complete the primary outcome assessment before continuing to complete the rest of the questionnaire. There were some cases where participants completed the primary outcome assessment but did not provide any data for the secondary outcome. Where no primary or secondary data points were provided, these were excluded for the specific analysis.

Statistical Analysis

The statistical analysis plan was prospectively registered on the Open Science Framework [29]. Data were analyzed using Stata 16.1 MP (StataCorp).

Descriptive statistics were reported either as unweighted frequencies and percentages, the mean with the 95% CI, or the median with the IQR.

The primary and secondary outcomes were modeled using linear mixed effects models. Each outcome was tested in a separate model. Each model included up to 3 repeated outcome assessments, collected at days 28, 84, and 168. Repeated measures were clustered within individual participants, represented with a random intercept. The mixed model used all available information (ie, participants with at least 1 follow-up assessment were analyzed), leading to more precise estimates of the treatment effect.

We used multivariable binary logistic regression to assess whether baseline variables were associated with missingness in the primary outcome variable (1=missing primary outcome at day 84, 0=nonmissing). Each model included as covariates (1) time (measured as days since baseline), (2) a dummy variable to represent treatment allocation (0=control, 1=intervention), (3) a time × arm interaction term, (4) the baseline measurement of the outcome, (5) relevant covariates (age and gender), and

(6) baseline variables associated with missingness (number of days off work due to alcohol consumption). Treatment effects were estimated as the difference between baseline and follow-up assessment (day 28, 84, or 168) and reported as the absolute alcohol unit difference between the arms.

Between-group effect sizes (Cohen *d*) were calculated by subtracting baseline total units consumed/AUDIT score from the day-of-assessment margin mean values (d=0.2, small effect; d=0.5, intermediate effect; and d=0.8, strong effect). The threshold for statistical significance reported in these analyses was *P*=.05.

The intention-to-treat analyses included all participants who completed at least 1 follow-up assessment (day 28, 84, or 168).

Sensitivity Analysis

We conducted a predefined sensitivity analysis in a subgroup of participants who had complete information for the primary outcome at day 84 (complete case analysis).

Process Evaluation

We examined process evaluation measures, used as a proxy for app usage. These were reported in 3 categories: (1) app utilization based on app analytics data provided by Google Analytics, (2) drinking analytics based on server interactions, and (3) notifications sent by the server. Where appropriate, these were reported either as the median with the IQR or as the mean with SD.

Usability

We examined usability of the *Drinks*:Ration app using the mobile health (mHealth) App Usability Questionnaire (MAUQ) [28] at day 28. Questionnaire responses were aggregated into (1) overall usability, (2) ease of use, (3) interface and satisfaction, and (4) usefulness. Results were summarized with means and SDs.

The study was also reported following the Template for Intervention Description and Replication [30] and the CONSORT (Consolidated Standards of Reporting Trials [31] and eHealth version [32]) checklist.

Results

Study Participation, Sample Characteristics, and Attrition

Between October 2020 and April 2021, 2708 individuals were invited to take part, of whom 2531 (93.5%) did not respond to the invite or declined to take part (n=150, 5.5%). In total, 177 (6.5%) participants were invited to complete a baseline assessment, of whom 54 (30.5%) were found to be ineligible based on study criteria (Figure 2). Therefore, a total of 123 (4.5%) participants completed the baseline assessment and were randomized into the study. Of these, 78 (63.4%) completed outcome assessments at day 28, 79 (64.2%) completed outcome assessments at day 84, and 27 (22.0%) completed outcome assessments at day 168. A total of 19 (15.4%) participants withdrew from the study by day 84. This included 7 (36.8%) participants who withdrew due to the limited functionality of the control version of the app.

Of the 123 participants, 62 (50.4%) participants were randomized to the intervention arm and 61 (49.6%) to the control arm. Baseline characteristics are shown in Table 2. The overall mean age was 47.6 years (95% CI 45.8-49.3), 117 (95.1%) participants were male, and 95 (77.2%) were married or in a long-term relationship. In addition, 87 (70.7%) had served in the army, and on average, the participants had served 14.4 years (95% CI 12.9-15.9) in the UK AF. The participants had a median AUDIT score of 16 (IQR 10-22) at baseline, and 66 (53.7%) were identified as having no probable PTSD. A total of 65 (52.9%) participants reported probable depression. Most participants entered the study with an Android device (n=67, 54.5%), and 79 (64.3%) participants completed the primary outcome assessment at day 84, with 76 (61.8%) completing the secondary outcome assessment.

Figure 2. CONSORT diagram for recruitment into the RCT. CONSORT: Consolidated Standards of Reporting Trials; RCT: randomized controlled trial.

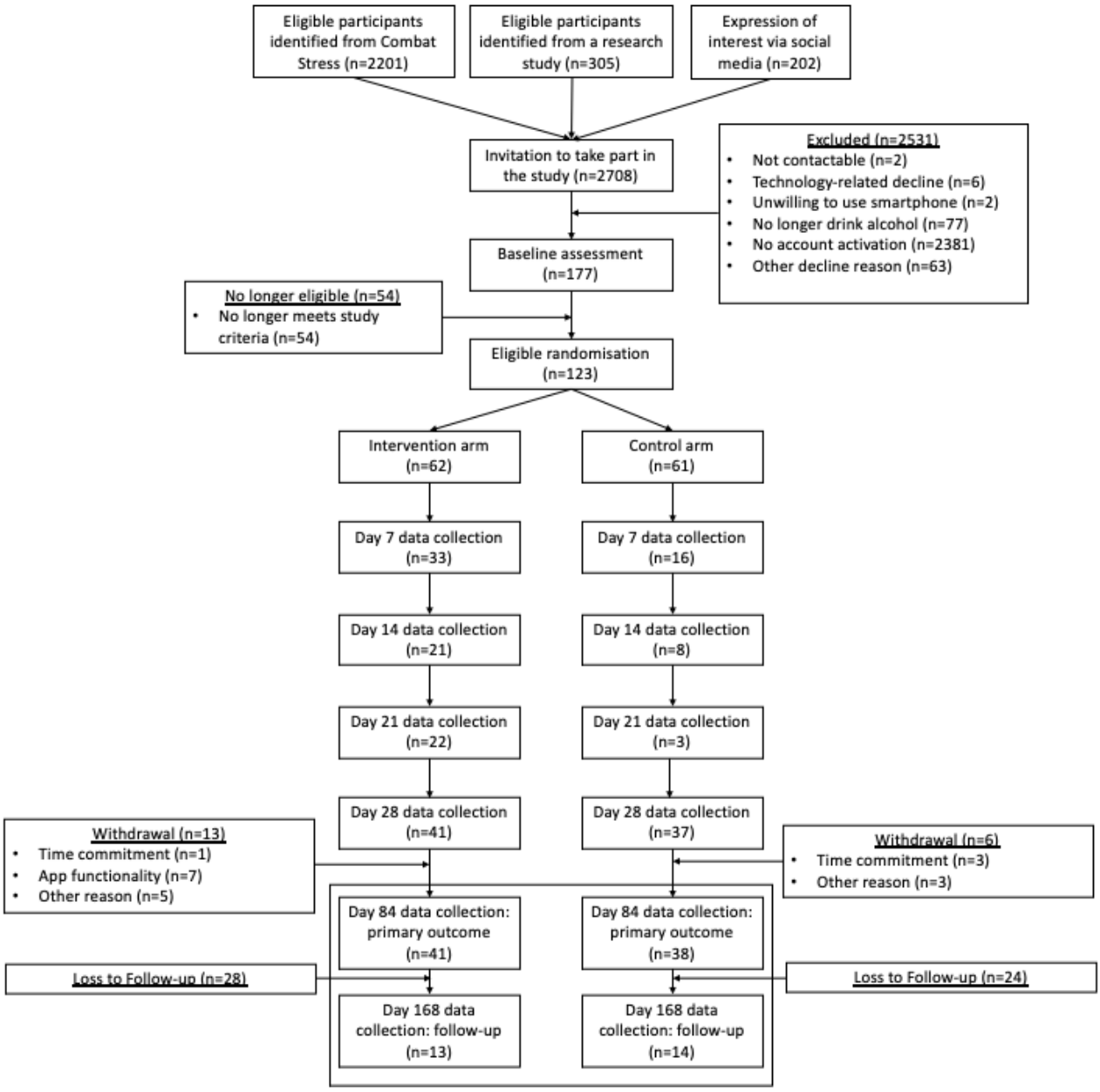


Table 2. Descriptive statistics of eligible participants.

Characteristics	Total (N=123)	Control (N=61)	Intervention (N=62)
Age (years), mean (95% CI)	47.6 (45.8-49.3)	47.4 (44.9-50.0)	47.7 (45.3-50.1)
Gender, n (%)			
Male	117 (95.1)	58 (95.1)	59 (95.2)
Female	6 (4.9)	3 (4.9)	3 (4.8)
Marital status, n (%)			
Married/in a relationship	95 (77.2)	48 (78.7)	47 (75.8)
Single/separated	19 (15.5)	8 (13.1)	11 (17.7)
Divorced/widowed	9 (7.3)	5 (8.2)	4 (6.5)
Military branch, n (%)			
Royal Navy/Royal Marines	14 (11.4)	6 (9.8)	8 (12.9)
Army	87 (70.7)	44 (73.2)	43 (69.6)
Royal Air Force	16 (13.1)	8 (13.1)	8 (13.1)
Other[a]	6 (4.9)	3 (4.9)	3 (4.9)
Length of military service in years, mean (95% CI)	14.4 (12.9-15.9)	15 (12.9-17.1)	13.8 (11.5-16.1)
Probable PTSD[b], n (%)	57 (46.3)	26 (42.6)	31 (50.0)
Probable depression, n (%)	65 (52.9)	30 (49.2)	35 (56.5)
Probable anxiety, n (%)	61 (49.6)	32 (52.5)	29 (46.8)
AUDIT[c] score, median (IQR)	16 (10-22)	14 (8-23)	16 (12-21)
Baseline unit weekly alcohol consumption[d], median (IQR)	44 (25-70)	43 (25-62)	47 (26-73)
Device type, n (%)			
iOS	56 (45.5)	32 (52.5)	24 (38.7)
Android	67 (54.5)	29 (47.5)	38 (61.3)
Withdrawn by day 84, n (%)	19 (15.5)	13 (21.3)	6 (9.7)
Completed primary outcome assessment, n (%)			
Day 28	78 (63.4)	37 (60.7)	41 (66.1)
Day 84	79 (64.3)	38 (62.3)	41 (66.1)
Day 168	27 (22.0)	14 (23.0)	13 (21.0)
Completed secondary outcome assessment, n (%)			
Day 28	73 (59.4)	34 (55.7)	39 (62.9)
Day 84	76 (61.8)	37 (60.7)	39 (62.9)
Day 168	27 (22.0)	14 (23.0)	13 (21.0)

[a]Service branch not reported in medical records.

[b]PTSD: posttraumatic stress disorder.

[c]AUDIT: Alcohol Use Disorder Identification Test (10 items).

[d]Recorded using the Timeline Follow Back for Alcohol Consumption (TLFB).

Primary and Secondary Outcome Analysis

For the primary outcome of the TLFB (units of alcohol per week) at day 84 (Table 3), we found that participants in the intervention arm had significantly larger reductions in self-reported alcohol unit consumption from baseline (marginal unit mean 56.3, 95% CI 50.6-62.0) to day 84 (marginal unit mean 28.1, 95% CI 21.1-35.1) compared with those in the control arm (marginal unit mean from 54.0, 95% CI 48.2-59.8, to 43.5, 95% CI 36.3-50.8; interaction *P*=.01). The effect size for the difference between the intervention and control arms in the mean change of units between baseline and day 84 was Cohen *d*=0.35, which is consistent with a moderate effect size.

Overall, we found that between baseline and day 84, weekly alcohol consumption reduced by –10.5 (95% CI –19.5 to –1.5) units in the control arm and –28.2 (95% CI –36.9 to –19.5) units in the intervention arm (*P*-value for the difference between arms

at day 84=.003) at the primary outcome measure. The difference in unit marginal means was –15.4 (95% CI –25.5 to –5.4) units of alcohol in favor of the intervention arm.

There was evidence of a strong effect between the 2 arms by day 28 (Cohen *d*=0.50) but no evidence of a difference between the 2 arms by day 168 (Cohen *d*=0.11) for self-reported alcohol consumption (Figure 3).

For the secondary outcome change in the AUDIT score by day 84, we again found that participants in the intervention arm had significantly larger reductions in the AUDIT score from baseline (score marginal mean 16.3, 95% CI 15.0-17.5) to day 84 (score marginal mean 10.1, 95% CI 8.5-11.8) compared with those in the control arm (score marginal mean from 16.0, 95% CI 14.6-17.3, to 14.1, 95% CI 12.4-15.7; interaction *P*=.003). The difference was –3.9 (95% CI –6.2 to –1.6) points on the AUDIT score in favor of the intervention arm. The effect size for the difference between the intervention and control arms in the mean change in the AUDIT score between baseline and day 84 was Cohen *d*=0.48.

There was evidence of a strong effect between the 2 arms by day 28 (Cohen *d*=0.53) but no evidence of an effect between the 2 arms by day 168 (Cohen *d*=0.06) for the AUDIT score (Multimedia Appendix 3).

Sensitivity analyses of primary and secondary outcomes using complete case analysis produced the same patterns as those identified in the main analysis (Multimedia Appendix 4).

Table 3. Estimated mean change between each measure, timepoint, and arm. The difference in the rate of change between each arm compared with baseline is reported alongside the Cohen *d* statistic[a].

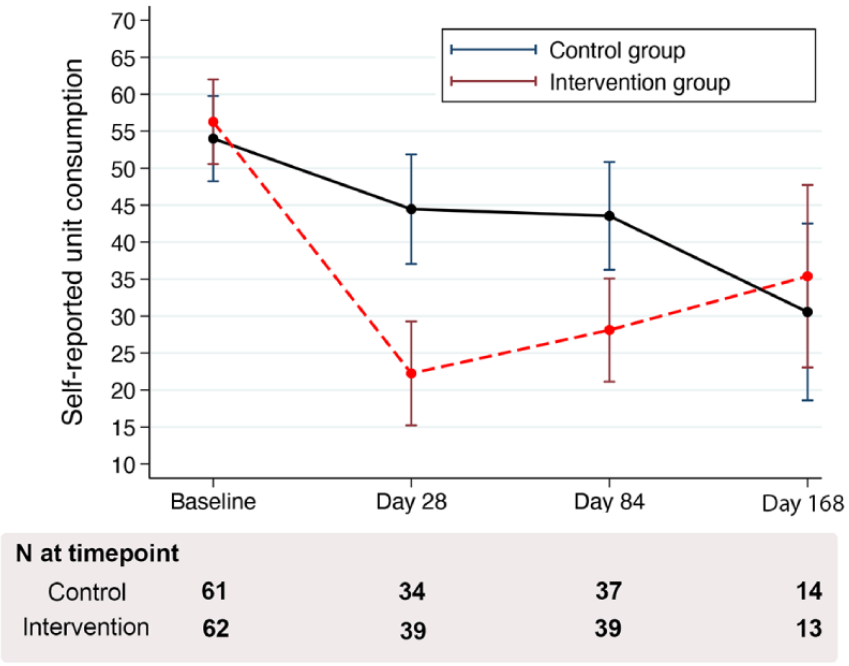
Study arm	Estimated marginal mean (95% CI)				Evidence for a difference in rate of change between arms, interaction *P* value			Cohen *d*		
	Baseline	Day 28	Day 84	Day 168	Baseline-day 28	Baseline-day 84 (primary outcome)	Baseline-day 168 (secondary outcome)	Baseline-day 28	Baseline-day 84 (primary outcome)	Baseline-day 168 (secondary outcome)
Self-reported units consumed over the previous week					<.001	.01	.80	0.50	0.35	0.11
Control	54.0 (48.2-59.8)	44.5 (37.1-51.9)	43.5 (36.3-50.8)	30.6 (18.6-42.5)	N/A[b]	N/A	N/A	N/A	N/A	N/A
Intervention	56.3 (50.6-62.0)	22.2 (15.2-29.3)	28.1 (21.1-35.1)	35.4 (23.1-47.7)	N/A	N/A	N/A	N/A	N/A	N/A
AUDIT[c] 10 score					.001	.003	.68	0.53	0.48	0.07
Control	16.0 (14.6-17.3)	16.5 (14.7-18.2)	14.1 (12.4-15.7)	13.2 (10.6-15.9)	N/A	N/A	N/A	N/A	N/A	N/A
Intervention	16.3 (15.0-17.5)	12.1 (10.5-13.7)	10.1 (8.5-11.8)	12.7 (9.9-15.4)	N/A	N/A	N/A	N/A	N/A	N/A

[a]Derived from a model that was adjusted for age, gender, number of days off work due to alcohol consumption, and outcome measure.

[b]N/A: not appliable.

[c]AUDIT: Alcohol Use Disorder Identification Test (10 items).

Figure 3. Trajectory for self-reported alcohol unit consumption per week as estimated from the mixed model.



Process Evaluation

Over the entire study period (168 days), participants in the control arm used the app for a median of 1 week (IQR 1-2), initialized the app a median of 3 times (IQR 2-9), and had a median session duration of 60.9 seconds (IQR 35.7-75.6). Participants in the intervention arm used the app for a median of 3.5 weeks (IQR 2-6), initialized the app a median of 13.5 times (IQR 4-27), and had a median session duration of 43.8 seconds (IQR 32.3-67.9); see Table 4.

Table 4. Engagement and app interactions over the study period per participant stratified by arm.

Interactions	Control, median (IQR)	Intervention, median (IQR)
Engagement measures		
Initializations	3 (2-9)	13.5 (4-27)
Session count	24 (16-45)	54 (27-150)
Session duration	60.9 (35.7-75.6)	43.8 (32.3-67.9)
Server interactions	7 (5-8)	13 (8-19)
App-recorded interactions		
Drinking days	N/A[a]	7 (4-11)
Drink-free days	N/A	3.5 (2-7)
Units consumed per drinking day	N/A	12.8 (4.4-16.5)
Notifications		
Push notifications	1 (1-1)	18 (9-19)
SMS[b] text messages	2 (0-2)	12 (10-14)
Weeks active	1 (1-2)	3.5 (2-6)

[a]N/A: not applicable; participants in the control arm were not able to provide this information.

[b]SMS: short message service.

Participants in the intervention arm reported a median of 7 drinking days (IQR 4-11) during the first 28-day period, a median of 3.5 drink-free days (IQR 2-7), and a median of 12.8 units of alcohol per drinking day (IQR 4.4-16.5). A median of 18 push notifications (IQR 9-19) were sent to participants in the intervention arm, along with a median of 12 SMS text messages (IQR 10-14).

App use of participants in the intervention arm is shown in Table 5. Participants engaged with all modules of the app, but most of the app engagement was spent using the screening module (mean 201.0, SD 994.6) and the normative guidance module (mean 510.4, SD 1012.7).

Table 5. Intervention arm engagement with the *Drinks*:Ration app, stratified by page between baseline and day 168 based on app analytics data.

Page	Ever accessed[a]	Number of times accessed		Average time spent per session (seconds)
	n (%)	Mean (SD)	Median (IQR)	Median (IQR)
Screening	62 (100)	4.9 (4.9)	2 (2-6)	27.7 (6.6-121.7)
Normative Guidance	62 (100)	8.7 (9.9)	5.5 (3-10)	99.5 (34.5-459)
Consent	62 (100)	3.3 (0.9)	3.5 (3-4)	53.8 (44.7-68.0)
Dashboard	60 (96.8)	34.8 (60.5)	10.5 (4-42)	197.1 (85.0-360.8)
Add Drinks	55 (88.7)	30.1 (53.0)	11 (3-38)	197.7 (47.8-408.5)
Timeline Follow Back	52 (83.9)	3.1 (2.11)	2 (2-4)	1021.2 (700.1-1579.0)
Drinks Diary Information	52 (83.9)	17.1 (23.3)	11 (2-22)	626.0 (117.2-1413.6)
Drinks Diary	50 (80.7)	8.2 (13.1)	3 (1-8)	45.7 (16.5-198.8)
View Goals	47 (75.8)	3.3 (0.7)	2 (1-4)	31.9 (3.1-63.4)
User Account	40 (64.5)	3.0 (8.5)	1 (0-3)	17.3 (0-55.1)

[a]During the study, Apple (developer of the iOS operating system) changed policies related to how developers could track and monitor usage of an app. This required specific user content, which could be modified outside the app. It is therefore not possible to ascertain whether a user did not give data because they were not using the app or whether they declined to share app usage statistics.

Usability

The participants completed the MAUQ at day 28 (Table 6). They responded to a set of usability questions on a scale of 1-7, with a higher value indicating improved usability. Participants in the control arm reported a mean overall app usability score of 4.1 (SD 1.5), a mean ease-of-use score of 4.4 (SD 1.6), a mean interface and satisfaction score of 4.1 (SD 1.6), and a mean usefulness score of 3.6 (SD 1.7). These scores were lower than those of the intervention arm, which reported a mean overall app usability score of 5.9 (SD 1.1), a mean ease-of-use score of 5.9 (SD 1.2), a mean interface and satisfaction score of 5.9 (SD 1.1), and a mean usefulness score of 5.7 (SD 1.1).

Table 6. MAUQ[a] results at day 28, stratified by arm.

Items	Control (N=35), mean (SD)	Intervention (N=38), mean (SD)
Ease-of-use	4.4 (1.6)	5.9 (1.2)
Interface and satisfaction	4.1 (1.6)	5.9 (1.1)
Usefulness	3.6 (1.7)	5.7 (1.3)
Overall	4.1 (1.5)	5.9 (1.1)

[a]MAUQ: mobile health (mHealth) App Usability Questionnaire.

Adverse Events and Technical Issues

In total, 2 (1.6%) of 123 participants were identified as having a single adverse event of consuming more than 25 units of alcohol within 24 hours during the study period. Following our risk protocol [19], a signposting booklet to relevant charities was provided, as well as a call with the study clinical lead. After a clinical interview, both participants were allowed to continue in the study. Their treatment allocation was not disclosed to the participants. No other adverse events were identified. No technical issues occurred during the trial.

Discussion

Principal Findings

This study is 1 of the few RCTs to date to examine the efficacy of a fully automated 28-day brief alcohol intervention delivered via a mobile app in a help-seeking sample of UK veterans with at least hazardous alcohol consumption. Help-seeking veterans were consuming on average 55 units of alcohol per week at the outset of the study, well above the 14 units of alcohol per week recommended as the maximum by the UK Chief Medical Officer. At the primary outcome (day 84), the difference between the estimated marginal means for the intervention and control arms was 15.4 units of alcohol lower in the intervention arm than in the control arm. A similar pattern was also observed for the AUDIT score, where the difference between the estimated marginal mean between the arms was 4.0 points on the AUDIT scale, also lower than that of the control arm. Overall, the intervention arm achieved significantly better reductions in alcohol consumption and AUDIT score. These effects disappeared at day 168.

The findings of this RCT demonstrate the efficacy of an automated brief alcohol intervention with personalized messaging for those who consume alcohol at least at hazardous levels and have sought help for a mental health difficulty [15]. The findings also mirror those obtained in other studies [7,12]. In particular, the findings demonstrate the efficacy of *Drinks*:Ration within a group that has been shown to be at increased risk of dual diagnosis. The between-arm difference

compares favorably to stand-alone self-help interventions [33], and the differences are greater than those typically found with face-to-face therapies and the most successful therapist-guided interventions [34].

This study found significant differences between the arms for the primary and secondary outcomes at the primary data collection point (day 84), and the differences were even larger at day 28 when the first assessment took place. At follow-up (day 168), only 22% of participants responded, which limited our ability to discern differences at the timepoint. An alternative explanation is that the effect of the intervention reduced over time, so the long-term benefits of the *Drinks*:Ration app may need reinforcing beyond the intervention month (ie, the first 28 days).

The results of this RCT should be placed in the context of the wider literature. A recent literature review exploring the effectiveness of alcohol reduction apps and the availability of evidenced-based apps on top commercial app stores identified 21 articles representing 19 unique mobile apps [35]. Of these, 7 (36.8%) apps were targeted at adolescent drinkers, and the remainder on the general population. No studies that targeted the AF were identified. The overall effectiveness of the included interventions was mixed, with standards of reporting making direct comparisons difficult (eg, AUDIT-C, binge-drinking days, alcohol unit consumption). *Drinks*:Ration results compare favorably to all included studies, in so far as we identified the largest reductions in alcohol consumption; however, our base starting point was significantly higher than the general population.

Although there are no reported data on waiting times between referral and treatment in the United Kingdom, National Health Service (NHS) Scotland national drug and alcohol treatment waiting times for alcohol treatment were reported to be around 3 weeks or less between January 2021 and March 2021 [36]. The delay between referral and treatment may be an opportunity to deploy the *Drinks*:Ration app to support help-seeking veterans while they wait for formal treatment. Help-seeking veterans misusing alcohol attend fewer mental health appointments [6], probably because many are prevented from receiving treatment for mood disorders and PTSD until they have reduced their excessive drinking. The use of the *Drinks*:Ration app to support reduction in alcohol consumption could enable more help-seeking veterans to access services. In addition to the *Drinks*:Ration app, the use of personalized messages sent via SMS and push notification may have contributed to improved performance of the intervention arm.

The efficacy of the app in relation to our primary outcome is encouraging, but there are some issues worth considering. The recruitment for this study was lower than anticipated. We expected that 30% of eligible veterans would enter the study, but though it was difficult to estimate the true total number of eligible individuals, the percentage who consented to participate was less than 10%. The second potential problem is that the effect seen at 84 days may need a reinforcement to encourage persistent changes in behavior over the long term. This could be achieved by enabling personalized messaging over the entire life course of app usage.

There were 2 adverse events (involving 2 participants), which were unlikely to be caused by our app and more likely to be caused by the ongoing alcohol consumption of the individuals involved. This highlights the challenges of monitoring adverse events in remote/automated interventions for which the implementation of the app needs constant monitoring while being used. Finally, it is important to consider how *Drinks*:Ration can be integrated into the treatment pathway to support veterans prior and during treatment, while also monitoring the degree to which its efficacy transfers to a clinical context.

Limitations

Several limitations of this trial should be noted. First, as already acknowledged, the majority of those invited to participate in the study did not take part. It is not possible to ascertain why these individuals chose not to take part, but it may be due to digital fatigue because of the COVID-19 pandemic. Therefore, we consider that our study assessed the efficacy of the intervention in those willing to engage with the app rather than effectiveness in the target population. Second, participants self-identified their military and help-seeking status among those recruited through social media and the status was not verified. Third, we only used self-reported data provided via outcome assessments and did not use data collected via the drinks diary. This decision to use the TLFB for assessing our primary outcome was to ensure comparability with other studies and to also ensure the control arm did not complete the drinks diary. This resulted in duplicating participant input, which could have created user frustration and negatively impacted usability and participation. Finally, this RCT was conducted during the COVID-19 pandemic. This period resulted in meaningful behavioral changes to how UK military veterans consumed alcohol due to lockdown. In a recent study of UK veterans, they were found to be drinking less alcohol during the first phase of the pandemic, reducing their hazardous drinking from 49% to 28% [37]. This may have reduced the available population that consumes alcohol at a harmful-to-hazardous level.

Conclusion

Our findings suggest that *Drinks*:Ration is efficacious in reducing alcohol consumption in help-seeking veterans and that wider uptake of *Drinks*:Ration in this population would be beneficial. However, strategies to increase use of the app and ensure that the gains in decreasing alcohol consumption persist over time need to be well thought out. This could be achieved by promoting app use and continuation of messaging and more personalized goal setting.

Acknowledgments

This study was funded by the Forces in Mind Trust (Project FIMT/0323KCL), a funding scheme run by the Forces in Mind Trust using an endowment awarded by the National Lottery Community Fund. The views expressed are those of the authors and not necessarily those of the National Health Service (NHS), the National Institute for Health and Care Research (NIHR), or the Department of Health and Social Care.

Data Availability

Due to the sensitive nature of the cohort, access to data is limited. Researchers may apply to access a pseudonymized data set. Requests to access study data are subject to submission of a research proposal to the corresponding author. All requests must be made in accordance with the UK Policy Framework for Health and Social Care research. Where the applicant is outside of King's College London, a data-sharing agreement is required.

Authors' Contributions

DL, DM, NTF, and RR were responsible for conceptualization; DL, LG, DM, RR, NTF, AS, JS, and EC, methodology; CW and DM, data collection; DL, software; DL, JDD, and EC, data curation; EC and JDD, validation; DL, EC, and JDD, formal analysis; DL, writing—original draft; DL, CW, RR, EC, JS, JDD, AS, NT, LG, and DM, writing—review and editing; DL, DM, and LG, supervision; DL and CW, project administration; and DL, RR, LG, DM, and JS, funding acquisition.

Conflicts of Interest

NTF is partly funded by a grant from the UK Ministry of Defence. NTF sits on the Independent Group Advising on the Release of Patient Data at NHS Digital. NTF is also a trustee of a military-related charity. AS is a full-time member of the UK armed forces (AF) seconded to King's College London. DL is a reservist in the UK AF. DM is employed by Combat Stress, a national charity in the United Kingdom that provides clinical mental health services to veterans and is a trustee of the Forces in Mind Trust (the funder for the project). DL and EC are partly funded by the National Institute for Health and Care Research (NIHR) Maudsley Biomedical Research Centre at South London and the Maudsley NHS Foundation Trust and King's College London and represents independent research.

Multimedia Appendix 1
Messaging: push notifications and SMS texts. SME: short message service.
[DOCX File , 34 KB - mhealth_v10i6e38991_app1.docx]

Multimedia Appendix 2
Standard units.
[DOCX File , 30 KB - mhealth_v10i6e38991_app2.docx]

Multimedia Appendix 3
Trajectory for AUDIT outcome from the mixed models. AUDIT: Alcohol Use Disorder Identification Test.
[DOCX File , 65 KB - mhealth_v10i6e38991_app3.docx]

Multimedia Appendix 4
Complete case analysis.
[DOCX File , 33 KB - mhealth_v10i6e38991_app4.docx]

Multimedia Appendix 5
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 361 KB - mhealth_v10i6e38991_app5.pdf]

References

1. Rona RJ, Jones M, Fear NT, Hull L, Hotopf M, Wessely S. Alcohol misuse and functional impairment in the UK armed forces: a population-based study. Drug Alcohol Depend 2010 Apr 01;108(1-2):37-42. [doi: 10.1016/j.drugalcdep.2009.11.014] [Medline: 20047802]
2. Stevelink SAM, Jones M, Hull L, Pernet D, MacCrimmon S, Goodwin L, et al. Mental health outcomes at the end of the British involvement in the Iraq and Afghanistan conflicts: a cohort study. Br J Psychiatry 2018 Dec 08;213(6):690-697 [FREE Full text] [doi: 10.1192/bjp.2018.175] [Medline: 30295216]

3. Goodwin L, Norton S, Fear N, Jones M, Hull L, Wessely S, et al. Trajectories of alcohol use in the UK military and associations with mental health. Addict Behav 2017 Dec;75:130-137. [doi: 10.1016/j.addbeh.2017.07.010] [Medline: 28734152]
4. Murphy D, Turgoose D. Exploring patterns of alcohol misuse in treatment-seeking UK veterans: a cross-sectional study. Addict Behav 2019 May;92:14-19. [doi: 10.1016/j.addbeh.2018.11.044] [Medline: 30572207]
5. Murphy D, Ashwick R, Palmer E, Busuttil W. Describing the profile of a population of UK veterans seeking support for mental health difficulties. J Ment Health 2019 Dec 06;28(6):654-661. [doi: 10.1080/09638237.2017.1385739] [Medline: 28982279]
6. Elbogen EB, Wagner HR, Johnson SC, Kinneer P, Kang H, Vasterling JJ, et al. Are Iraq and Afghanistan veterans using mental health services? New data from a national random-sample survey. Psychiatr Serv 2013 Mar 01;64(2):134-141 [FREE Full text] [doi: 10.1176/appi.ps.004792011] [Medline: 23475498]
7. Fowler LA, Holt SL, Joshi D. Mobile technology-based interventions for adult users of alcohol: a systematic review of the literature. Addict Behav 2016 Nov;62:25-34. [doi: 10.1016/j.addbeh.2016.06.008] [Medline: 27310031]
8. Garnett C, Crane D, Michie S, West R, Brown J. Evaluating the effectiveness of a smartphone app to reduce excessive alcohol consumption: protocol for a factorial randomised control trial. BMC Public Health 2016 Jul 08;16(1):536-552 [FREE Full text] [doi: 10.1186/s12889-016-3140-8] [Medline: 27392430]
9. drinkaware. Want to Change the Way You Drink? You Can Do It with MyDrinkaware. URL: https://www.drinkaware.co.uk/ [accessed 2022-05-30] [WebCite Cache ID 6zZtHhyXH]
10. Michie S, Whittington C, Hamoudi Z, Zarnani F, Tober G, West R. Identification of behaviour change techniques to reduce excessive alcohol consumption. Addiction 2012 Aug;107(8):1431-1440. [doi: 10.1111/j.1360-0443.2012.03845.x] [Medline: 22340523]
11. One You Drink free days. URL: https://www.my-therappy.co.uk/app/drinks-tracker-1 [accessed 2022-05-30] [WebCite Cache ID 6zZtobwPF]
12. Kaner EF, Beyer FR, Garnett C, Crane D, Brown J, Muirhead C, et al. Personalised digital interventions for reducing hazardous and harmful alcohol consumption in community-dwelling populations. Cochrane Database Syst Rev 2017 Sep 25;9:CD011479 [FREE Full text] [doi: 10.1002/14651858.CD011479.pub2] [Medline: 28944453]
13. Williamson C, White K, Rona R, Simms A, Fear N, Goodwin L, et al. Smartphone-based alcohol interventions: a systematic review on the role of notifications in changing behaviours towards alcohol. Subst. Abuse 2022:1-17 (forthcoming).
14. Jones E, Fear NT. Alcohol use and misuse within the military: a review. Int Rev Psychiatry 2011 Apr 27;23(2):166-172. [doi: 10.3109/09540261.2010.550868] [Medline: 21521086]
15. Leightley D, Puddephatt J, Jones N, Mahmoodi T, Chui Z, Field M, et al. A smartphone app and personalized text messaging framework (InDEx) to monitor and reduce alcohol use in ex-serving personnel: development and feasibility study. JMIR Mhealth Uhealth 2018 Sep 11;6(9):e10074 [FREE Full text] [doi: 10.2196/10074] [Medline: 30206054]
16. Leightley D, Puddephatt J, Goodwin L, Rona R, Fear NT. InDEx: open source iOS and Android software for self-reporting and monitoring of alcohol consumption. J Open Res Softw 2018 Mar 23;6(1):13 [FREE Full text] [doi: 10.5334/jors.207] [Medline: 29795769]
17. Puddephatt J, Leightley D, Palmer L, Jones N, Mahmoodi T, Drummond C, et al. A qualitative evaluation of the acceptability of a tailored smartphone alcohol intervention for a military population: Information about Drinking for Ex-Serving personnel (InDEx) app. JMIR Mhealth Uhealth 2019 May 24;7(5):e12267 [FREE Full text] [doi: 10.2196/12267] [Medline: 31127726]
18. *Drinks*:Ration: an Android and iOS App for Armed Forces Personnel. URL: https://www.drinksration.app/ [accessed 2022-05-30]
19. Leightley D, Rona RJ, Shearer J, Williamson C, Gunasinghe C, Simms A, et al. Evaluating the efficacy of a mobile app (Drinks:Ration) and personalized text and push messaging to reduce alcohol consumption in a veteran population: protocol for a randomized controlled trial. JMIR Res Protoc 2020 Oct 02;9(10):e19720 [FREE Full text] [doi: 10.2196/19720] [Medline: 33006569]
20. Williamson C, Rona RJ, Simms A, Fear NT, Goodwin L, Murphy D, et al. Recruiting military veterans into alcohol misuse research: the role of social media and Facebook advertising. Telemed J E Health 2022 May 11 Ahead of print. [doi: 10.1089/tmj.2021.0587] [Medline: 35544055]
21. Sobell LC, Maisto SA, Sobell MB, Cooper A. Reliability of alcohol abusers' self-reports of drinking behavior. Behav Res Therapy 1979 Jan;17(2):157-160. [doi: 10.1016/0005-7967(79)90025-1]
22. Kunz FM, French MT, Bazargan-Hejazi S. Cost-effectiveness analysis of a brief intervention delivered to problem drinkers presenting at an inner-city hospital emergency department. J Stud Alcohol 2004 May;65(3):363-370. [doi: 10.15288/jsa.2004.65.363] [Medline: 15222593]
23. Hines LA, Goodwin L, Jones M, Hull L, Wessely S, Fear NT, et al. Factors affecting help seeking for mental health problems after deployment to Iraq and Afghanistan. Psychiatr Serv 2014 Jan 01;65(1):98-105. [doi: 10.1176/appi.ps.004972012] [Medline: 24037454]
24. Arroll B, Goodyear-Smith F, Crengle S, Gunn J, Kerse N, Fishman T, et al. Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. Ann Fam Med 2010 Jul 19;8(4):348-353 [FREE Full text] [doi: 10.1370/afm.1139] [Medline: 20644190]

25. Kroenke K, Spitzer RL, Williams JB, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med 2007 Mar 06;146(5):317. [doi: 10.7326/0003-4819-146-5-200703060-00004]
26. Cloitre M, Shevlin M, Brewin CR, Bisson JI, Roberts NP, Maercker A, et al. The International Trauma Questionnaire: development of a self-report measure of ICD-11 PTSD and complex PTSD. Acta Psychiatr Scand 2018 Dec 03;138(6):536-546. [doi: 10.1111/acps.12956] [Medline: 30178492]
27. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II. Addiction 1993 Jun;88(6):791-804. [doi: 10.1111/j.1360-0443.1993.tb02093.x] [Medline: 8329970]
28. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500 [FREE Full text] [doi: 10.2196/11500] [Medline: 30973342]
29. Leightley D, Rona RJ, Shearer J, Williamson C, Gunasinghe C, Simms A, et al. Evaluating the Efficacy of a Mobile App (Drinks:Ration) and Personalised Text and Push Messaging to Reduce Alcohol Consumption in a Veteran Population: Randomised Controlled Trial. Open Science Framework 2022 May 28 [FREE Full text] [doi: 10.17605/OSF.IO/QHPBZ]
30. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ 2014 Mar 07;2014(348):g1687. [doi: 10.1136/bmj.g1687] [Medline: 24609605]
31. Rennie D. CONSORT Revised: improving the reporting of randomized trials. JAMA 2001 Apr 18;285(15):2006-2007. [doi: 10.1001/jama.285.15.2006] [Medline: 11308440]
32. 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]
33. Riper H, Blankers M, Hadiwijaya H, Cunningham J, Clarke S, Wiers R, et al. Effectiveness of guided and unguided low-intensity internet interventions for adult alcohol misuse: a meta-analysis. PLoS One 2014 Jun;9(6):e99912 [FREE Full text] [doi: 10.1371/journal.pone.0099912] [Medline: 24937483]
34. Doherty A, Mason C, Fear N, Rona R, Greenberg N, Goodwin L. Are brief alcohol interventions targeting alcohol use efficacious in military and veteran populations? A meta-analysis. Drug Alcohol Depend 2017 Sep 01;178:571-578. [doi: 10.1016/j.drugalcdep.2017.05.029] [Medline: 28750345]
35. Colbert S, Thornton L, Richmond R. Smartphone apps for managing alcohol consumption: a literature review. Addict Sci Clin Pract 2020 May 07;15(1):17 [FREE Full text] [doi: 10.1186/s13722-020-00190-x] [Medline: 32381062]
36. Public Health Scotland. National Drug and Alcohol Treatment Waiting Times - 1 January to 31 March 2021. URL: https://publichealthscotland.scot/publications/national-drug-and-alcohol-treatment-waiting-times/national-drug-and-alcohol-treatment-waiting-times-1-january-to-31-march-2021/ [accessed 2021-12-23]
37. Sharp M, Serfioti D, Jones M, Burdett H, Pernet D, Hull L, et al. UK veterans' mental health and well-being before and during the COVID-19 pandemic: a longitudinal cohort study. BMJ Open 2021 Aug 27;11(8):e049815 [FREE Full text] [doi: 10.1136/bmjopen-2021-049815] [Medline: 34452965]

Abbreviations

AF: armed forces
AUDIT: Alcohol Use Disorder Identification Test
BCT: behavior change technique
InDEx: Information about Drinking for Ex-Serving personnel
MAUQ: mobile health (mHealth) App Usability Questionnaire
PTSD: posttraumatic stress disorder
RCT: randomized controlled trial
SMS: short message service
TLFB: Timeline Follow Back for Alcohol Consumption

Edited by L Buis; submitted 25.04.22; peer-reviewed by M Almond, M Bardus; comments to author 10.05.22; revised version received 12.05.22; accepted 19.05.22; published 20.06.22.

Please cite as:
Leightley D, Williamson C, Rona RJ, Carr E, Shearer J, Davis JP, Simms A, Fear NT, Goodwin L, Murphy D
Evaluating the Efficacy of the Drinks:Ration Mobile App to Reduce Alcohol Consumption in a Help-Seeking Military Veteran Population: Randomized Controlled Trial
JMIR Mhealth Uhealth 2022;10(6):e38991
URL: https://mhealth.jmir.org/2022/6/e38991
doi:10.2196/38991
PMID:35724966

Original Paper

Effectiveness of a Conversational Chatbot (Dejal@bot) for the Adult Population to Quit Smoking: Pragmatic, Multicenter, Controlled, Randomized Clinical Trial in Primary Care

Eduardo Olano-Espinosa[1*], MD, PhD; Jose Francisco Avila-Tomas[2,3*], MD; Cesar Minue-Lorenzo[4*], MD; Blanca Matilla-Pardo[5*], MD; María Encarnación Serrano Serrano[6*], MD, PhD; F Javier Martinez-Suberviola[7*], MD, PhD; Mario Gil-Conesa[8*], MD; Isabel Del Cura-González[9,10*], MD, PhD; Dejal@ Group[11*]

[1]Healthcare Center Los Castillos, Madrid Health Service, Alcorcón, Spain

[2]Healthcare Center Santa Isabel, Madrid Health Service, Leganes, Spain

[3]Preventive Medicine and Public Health Department, Rey Juan Carlos University, Alcorcon, Spain

[4]Health Care Center Perales del Rio, Madrid Health Service, Getafe, Spain

[5]Health Care Center Panaderas, Madrid Health Service, Fuenlabrada, Spain

[6]Healthcare Center Los Fresnos, Madrid Health Service, Torrejón de Ardoz, Spain

[7]Healthcare Center Guayaba, Madrid Health Service, Madrid, Spain

[8]Preventive Medicine Service, Hospital Universitario Fundación Alcorcón, Madrid Health Service, Madrid, Spain

[9]Research Unit, Primary Care Assistance Management, Madrid Health Service, Madrid, Spain

[10]Research Network on Health Services in Chronic Diseases, Instituto de Salud Carlos III (ISCIII), Madrid, Spain

[11]See Acknowledgments

[*]all authors contributed equally

Corresponding Author:
Eduardo Olano-Espinosa, MD, PhD
Healthcare Center Los Castillos
Madrid Health Service
Calle Carballino 25
Alcorcón, 28924
Spain
Phone: 34 679325778
Email: e.oeoeoeoe@gmail.com

Abstract

Background: Tobacco addiction is the leading cause of preventable morbidity and mortality worldwide, but only 1 in 20 cessation attempts is supervised by a health professional. The potential advantages of mobile health (mHealth) can circumvent this problem and facilitate tobacco cessation interventions for public health systems. Given its easy scalability to large populations and great potential, chatbots are a potentially useful complement to usual treatment.

Objective: This study aims to assess the effectiveness of an evidence-based intervention to quit smoking via a chatbot in smartphones compared with usual clinical practice in primary care.

Methods: This is a pragmatic, multicenter, controlled, and randomized clinical trial involving 34 primary health care centers within the Madrid Health Service (Spain). Smokers over the age of 18 years who attended on-site consultation and accepted help to quit tobacco were recruited by their doctor or nurse and randomly allocated to receive usual care (control group [CG]) or an evidence-based chatbot intervention (intervention group [IG]). The interventions in both arms were based on the 5A's (ie, Ask, Advise, Assess, Assist, and Arrange) in the US Clinical Practice Guideline, which combines behavioral and pharmacological treatments and is structured in several follow-up appointments. The primary outcome was continuous abstinence from smoking that was biochemically validated after 6 months by the collaborators. The outcome analysis was blinded to allocation of patients, although participants were unblinded to group assignment. An intention-to-treat analysis, using the baseline-observation-carried-forward approach for missing data, and logistic regression models with robust estimators were employed for assessing the primary outcomes.

Results: The trial was conducted between October 1, 2018, and March 31, 2019. The sample included 513 patients (242 in the IG and 271 in the CG), with an average age of 49.8 (SD 10.82) years and gender ratio of 59.3% (304/513) women and 40.7% (209/513) men. Of them, 232 patients (45.2%) completed the follow-up, 104/242 (42.9%) in the IG and 128/271 (47.2%) in the CG. In the intention-to-treat analysis, the biochemically validated abstinence rate at 6 months was higher in the IG (63/242, 26%) compared with that in the CG (51/271, 18.8%; odds ratio 1.52, 95% CI 1.00-2.31; P=.05). After adjusting for basal CO-oximetry and bupropion intake, no substantial changes were observed (odds ratio 1.52, 95% CI 0.99-2.33; P=.05; pseudo-R^2=0.045). In the IG, 61.2% (148/242) of users accessed the chatbot, average chatbot-patient interaction time was 121 (95% CI 121.1-140.0) minutes, and average number of contacts was 45.56 (SD 36.32).

Conclusions: A treatment including a chatbot for helping with tobacco cessation was more effective than usual clinical practice in primary care. However, this outcome was at the limit of statistical significance, and therefore these promising results must be interpreted with caution.

Trial Registration: Clinicaltrials.gov NCT 03445507; https://tinyurl.com/mrnfcmtd

International Registered Report Identifier (IRRID): RR2-10.1186/s12911-019-0972-z

(JMIR Mhealth Uhealth 2022;10(6):e34273) doi:10.2196/34273

KEYWORDS

smoking; tobacco cessation; primary care; smartphone use; chatbot; dialog systems; artificial intelligence; tobacco; mHealth; primary care

Introduction

Tobacco addiction is the leading cause of preventable morbidity and mortality in the world, directly causing 7 million deaths annually. Should this trend continue, this figure would rise to 8 million deaths by 2030, mostly in developing countries [1].

Population studies repeatedly conclude that the majority of smokers would like to quit and the percentage of them who try every year is high [1]. Most tobacco users stop smoking without help, although professional interventions increase the number of attempts and use of effective medication, resulting in a 2- to 3-fold success rate in the long term [2].

Different interventions by health professionals have proven to be effective and efficient, with the best outcomes observed when behavioral and pharmacological treatments are combined [2,3]. However, only 1 in 20 cessation attempts is supervised by a health professional [3]. Almost 84% of smokers who attended a primary care within the Madrid Health Service in 2008 had not received any advice to quit smoking over the 3 months prior to the consultation [4], which is similar to reports from other countries [5,6]. Factors accounting for these low intervention rates have been identified, among which are the training deficit of professionals, their perception that these interventions are not very effective, and their lack of time to implement them [7].

More intensive clinical interventions yield higher cessation rates in the long term; however, they are more expensive, require specifically trained professionals, and entail more health care time, which are inconvenient for both health care providers and users, who occasionally cannot afford them [8]. The potential advantages of mobile technologies for health (mHealth) [9]—effectiveness, accessibility, portability, privacy, customization, time-sensitive interventions, access to social support, superior adherence, and enormous scalability potential—can circumvent these problems and facilitate tobacco-cessation interventions for public health systems.

Globally, the number of smartphones used is increasing. There are an estimated 5200 million cell phone users and an estimated 8 billion cell phone lines worldwide, which exceed the world population (penetration rate of 102%) [10], and these numbers are expected to continue rising. Smartphones have become the most frequent and most accessible form of computer in most countries. This relevance of information and communication technologies (ICTs) has even increased in the context of the COVID-19 pandemic due to the imposed social distancing, and tobacco addiction was not oblivious to the new circumstances.

Using ICTs also entails risks: access to websites and apps offering incomplete information or nonevidence-based therapies that are difficult to identify and can cause undesirable effects [11]; incorrect records due to anonymity of patients (including the difficulty to reach the target population) [12]; lack of nonverbal communication; potential discrimination (eg, impaired vision, illiteracy, socioeconomic level, age); feeling of invasion of privacy or of being controlled for the user; issues with adherence to treatment and its detection; costs generated from mobile data use; and problems regarding data protection, privacy, and confidentiality. Online interventions should complement and not substitute presential interventions for now [13], so creating a theoretical frame for correctly implementing this novel type of interventions is essential to guarantee minimum quality and homogeneity standards.

Evidence regarding the effectiveness of interventions for quitting smoking with the aid of ICTs is recent. A review by Whittaker et al [14], which included 26 clinical trials and 33,849 participants, concluded that automatized interventions with SMS text messages were effective, whether as the solely delivered intervention (relative risk [RR] 1.54, 95% CI 1.19-2.00) or in combination with other interventions (RR 1.59, 95% CI 1.09-2.33). That review was the first to incorporate 5 evidence-based, quality studies comparing the effectiveness of an app for cessation with low-intensity interventions (whether using apps or not), although the effectiveness of apps for increasing the abstinence rates in the long term was not proven

(RR 1.00, 95% CI 0.66-1.52). A more recent review including 4 trials using apps reported similar results (RR 0.871, 95% CI 0.543-1.397) [15].

Chatbots are potentially useful tools for interventions using ICTs: they are virtual assistants that respond to questions and requests by the patients, have the ability to learn, and communicate with the user via messaging apps. They differ from apps in their structure (they do not require installation in the smartphone, and therefore do not occupy space in the terminal; the interface is like a chatroom; and programming-related costs and time are lower), usage (they are bidirectional communication tools), interaction (they are not limited to a series of actions set by the programmer), privacy (they do not collect data from the phone), and most importantly, they are artificial intelligence (AI) and natural language processing tools (Multimedia Appendix 1).

At the time of this writing, several clinical trials are being conducted to assess the effectiveness of a chatbot for quitting smoking [16,17] by comparing different interventions employing ICTs. However, this work considered that clinical practice was the best comparator because it is the standard treatment in our setting and the chatbot aims to reproduce the ideal professional-patient personalized interaction using novel technological support.

Given its easy scalability to large populations, chatbots are a potentially useful complement to usual treatment, with the consequent savings, whether they are integrated into a global plan for aiding smokers to quit or used alone.

The aim of this study was to assess the effectiveness of a chatbot, with an evidence-based design and including elements of AI and natural language processing, for helping people stop smoking compared with clinical practice in primary care.

Methods

Trial Design

This is a pragmatic, multicenter, controlled, and randomized clinical trial. The study was conducted in 34 primary health care centers in the Community of Madrid region (Spain) and had a follow-up period of 6 months. The Madrid Health Service provided care for 6,772,465 citizens in 262 health care centers when the trial was conducted (2019).

The study followed the CONSORT (Consolidated Standards of Reporting Trials) guidelines [18] (Multimedia Appendix 2). The trial protocol was previously registered [19] and no changes were made to the methods, intervention, or comparator, except for an additional analysis by subgroups, which was not included in the initial study design.

Participants

Family practitioners and nurses from the 262 health care centers in the Madrid Health Service were offered to participate. The 248 health workers who volunteered as collaborators were informed of the study objectives, design, and methods, and received training about the fieldwork, handling of the data collection, and good practice in clinical research. Among them, only 161 professionals recruited participants.

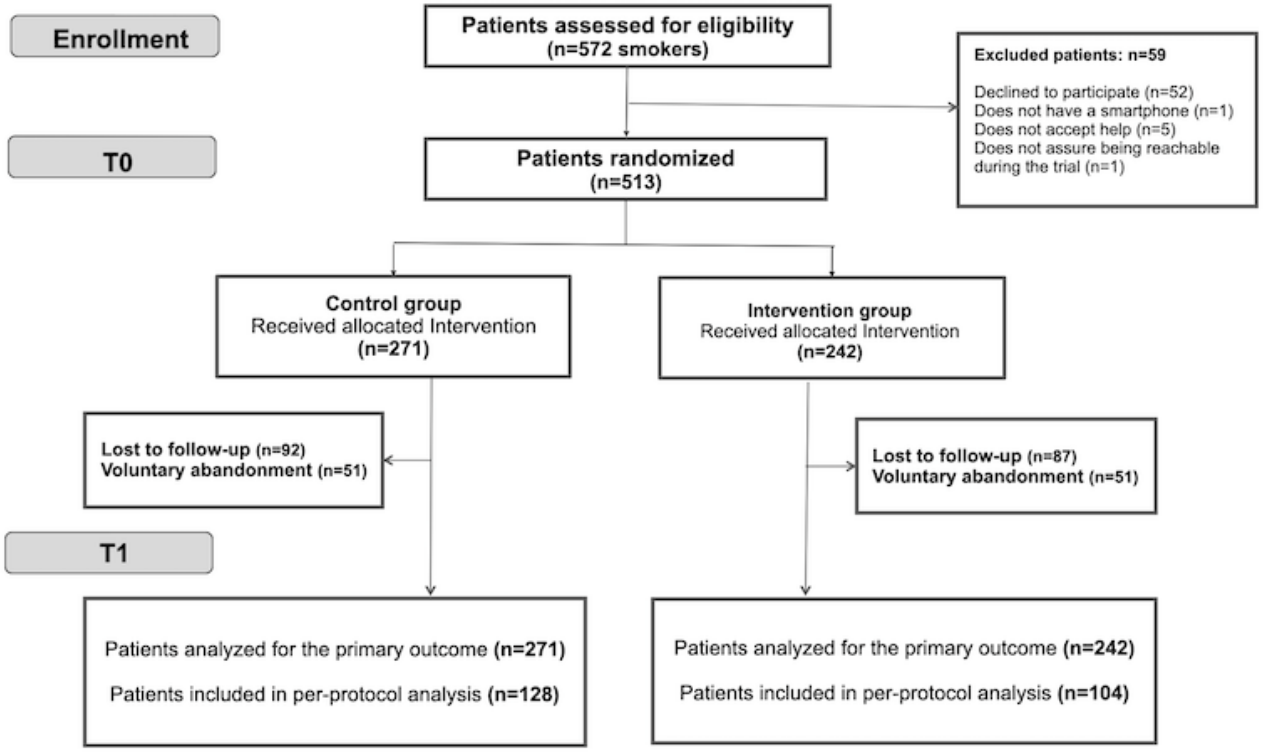
Patients included were smokers aged over 18 years who visited their doctor or nurse for consultation for any reason during the inclusion period. Patients included must have smoked at least one cigarette over the previous month, accept professional help for quitting in the following month, own a smartphone in which a messaging app (Telegram) could be installed, confirm their availability to be reached for 6 months following the intervention, and provide informed written consent. Criteria for exclusion were showing significant communication barriers and participation in another dishabituation program or clinical trial simultaneously. Computer or internet illiteracy of patients was not assessed.

Recruitment

Each collaborator had the objective of recruiting a minimum of 3 patients by offering participation to all smokers attending their consultation for any reason, in consecutive order, between October 1, 2018, and March 31, 2019. After checking compliance with the inclusion criteria, the patients were informed about the characteristics of the trial, and invited to participate and read an informative document (Multimedia Appendix 3). Patients who accepted to participate provided informed consent (Multimedia Appendix 4). Relevant data on patients who declined enrollment were collected (age, gender, and reason for declining).

Two visits were defined for patient data collection (Figure 1): baseline (T0) and at 6 months (T1). The health care collaborators collected participants' data in a collection notebook designed ad hoc, which could be accessed from the work computer with a personal password. Additionally, professionals were responsible for the clinical follow-up of patients in the control group (CG) and keeping records of it at each visit.

Figure 1. Study Flowchart.



Randomization and Blinding

After providing informed consent and following the collection of baseline information, participants were randomly allocated to the intervention group (IG, chatbot) or CG (usual care) at the baseline visit (T0) via simple randomization software and without further restrictions. No other method was used to implement the random allocation sequence. The software generated a banner indicating the professional which group the patient had been assigned to, and a printable file with a password to access the chatbot for patients in the IG.

Given the nature of the intervention, patients and professionals were aware of their treatment allocation. All analyses were performed by trial statisticians and methodologists in the Madrid Primary Care Research Unit who were blinded to the group assignment.

Intervention

The intervention strategy for both arms was based on the 5A's (ie, Ask, Advise, Assess, Assist, and Arrange) in the US Clinical Practice Guideline [2]. During the recruitment phase, all patients who met the inclusion criteria were interviewed in person about their tobacco consumption and received advice to cease smoking from their doctor or nurse, who also inquired about their willingness to quit smoking. Those accepting to attempt cessation in the following month and agreed to participate in the trial were randomly assigned into the IG or the CG. Patients received a personal intervention that combined behavioral and pharmacological treatment and was structured in several follow-up visits, whether online via a chatbot or face-to-face with their assigned health care professional.

Patients in the CG received usual clinical practice that aided in their tobacco cessation process, which is based on scientific recommendations and protocols in the services portfolio of the Madrid Health System (Servicio 415, Atención al Consumo de Tabaco en el Adulto). The standard intervention comprised at least one appointment before the day of cessation and another visit after 1 month. Additional controls could be set, with frequency, intensity, and duration adjusted based on the professional criteria and patient needs.

Patients in the IG were offered an intervention with contents similar to the CG but delivered via a chatbot. A personal keyword allowed accessing the chatbot via Telegram, a widely used messaging app very similar to others, which makes it very easy to use and was chosen due to its better privacy warranties at the moment this trial was conducted. No instructions or recommendations were given to the chatbot users regarding timing, frequency, or intensity of use. No further appointments were set between the professional and the patient other than the follow-up at 6 months (T1), and no additional co-interventions were provided outside the trial setting.

Once the patients started the interaction upon their initiative, the chatbot guided them through the dishabituation process by establishing a 15-day period just before the cessation date with daily interactive contacts between the chatbot and patient. This was followed by encouraging and recognition messages by the chatbot after quitting that became more sporadic until completing the 6 months of abstinence. The contact frequency set by the chatbot varied depending on the time since quitting and patient characteristics (personal choice, type of tobacco use, personal risk situation, prescribed drug, abstinence-related

symptoms, and evolution). The patient could contact the chatbot at any time and place, and decided the duration and frequency of interactions. There were no payments in any case or direction. The only expenditure for patients was that derived from consumed mobile data.

Dejal@bot was developed based on scientific evidence by doctors with expertise in tobacco use and ICTs between 2015 and 2018 (see Multimedia Appendix 5 for screenshots of the app). Its internal structure is a script recreating the interaction between a professional and a patient that takes numerous variants as required by the patient's needs and characteristics. The chatbot is bidirectional and provides multimedia links to cessation advice (by providing access to evidence-based techniques with cognitive-behavioral, motivational, relapse-prevention, and problem-solving components); information about the prescribed medication for helping to quit; and advice on how to cope with abstinence-related problems and relaxation exercises in diverse formats, such as video, graphs, games, and web links (of note, all these are similar to the resources health care workers could offer to the patients in the CG). Dejal@bot also incorporates gamification elements (knowledge and skills acquisition) with a system for scoring points and obtaining badges that grant access to specific information depending on the abstinence period and personal needs. This feedback is complemented with messages of encouragement and emphasis on the achieved goals. The intervention was described in detail in the protocol [19] and in the TIDieR (Template for Intervention Description and Replication) checklist (Multimedia Appendix 6).

A pilot test was conducted prior to the beginning of the clinical trial to assess usability and to train the AI categories. The final version from the pilot study (February 2018) was implemented in the randomized controlled trial and its content was not modified at any stage.

No technical support service was available during the trial, which the authors believed would improve the chatbot accessibility and the retention rate.

Data collection was monitored weekly and the collaborators were contacted in case of incongruous or incomplete information.

The Dejal@bot structure is simple: (1) The user writes messages in the Telegram app installed into their smartphone; (2) Telegram anonymizes this message upon receipt by assigning an identification number to the user and forwards the message to the software installed in the research team server; (3) the software processes the message; (4) our reply is sent to Telegram; and (5) Telegram forwards the response to the user.

Telegram operates as a telephone service provider acting as a technological intermediate that does not store the content of the conversation. The chatbot only knows what the users say but there are no metadata in the conversation allowing their identification.

Dejal@bot works on an expert system that becomes more flexible in each decision by understanding the patients' needs through a probabilistic interpretation of their message using techniques based on Bayes' theorem. Therefore, the decision on which script to show next is not a prefixed sequence in a decision tree but rather works as follows: if the patient does not require special attention, the subsequent script is used in the order that has been preset in the expert system; however, if the patient requires special attention (eg, change of quit date, relapse, medication side effect), the Bayesian system detects this need and the specific script that has been preset for that particular case is used.

The AI layer has been generated using intelligent dictionaries of synonyms (48 classes with a total of 1127 terms), and therefore, the chatbot has learned different ways of expressing the same concept in natural language to respond appropriately regardless of the expression used by the user. The features and clinical content of the chatbot are presented in Textbox 1.

Textbox 1. Chatbot features and clinical content.

Features

- Installed outside the smartphone (instead installed on the research team server).
- Communicate with the user via messaging apps.
- Easy to learn.
- Bidirectional communication.
- Respects the privacy of the user.
- Artificial intelligence.
- Natural language processing.
- Structured in several follow-up visits.
- Participants contact the chatbot at any time.
- Participants decide the duration and frequency of interactions.
- Multimedia links.
- Gamification elements.

Clinical content

- Evidence-based techniques with cognitive-behavioral, motivational, relapse-prevention, and problem-solving components.
- Variable contact frequency depending on time since quitting and patient characteristics.
- Information about the prescribed medication.
- Advice on how to cope with abstinence-related problems.
- Relaxation exercises.
- 15-day period just before the cessation date with daily interactive contacts.
- Encouragement and recognition messages by the chatbot after quitting with an emphasis on the achieved goals.

Outcome Variables

The primary outcome was continuous abstinence at 6 months, which was biochemically validated by CO-oximetry, involving measurement of exhaled air in parts per million (ppm), following the recommendations in the Russell Standard [20]. Therefore, the patient must declare not having smoked in the previous 6 months and have a negative CO-oximetry result (<10 ppm) to be considered a "nonsmoker."

The secondary outcomes were changes in quality of life, number of contacts between the therapist or the chatbot and the patient, and total time of interaction. The cost-utility assessment will be the subject of a separate analysis and paper. Adherence to pharmacological treatment could not be measured due to the low number of visits in the CG. No modifications to the trial outcome measurements were made after the trial commenced.

The EQ-5D-5L was used to assess quality of life. This validated generic instrument measures health-related quality of life on a visual analog scale (VAS) ranging from 0 to 100 (with higher ratings indicating higher quality of life) and includes 5 dimensions to assess mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Health condition is converted into a weighted Health Status Index, where *full health* receives a value of 1 and 0 stands for *death*. We collected data on the 5 dimensions and the Health Status Index proposed for our country [21].

Patient data were collected during the consultations with family practitioners and nurses at baseline (T0) and after 6 months (T1). At baseline, collaborators recorded sociodemographic variables (age, gender, economical level, educational level, and nationality); tobacco use (daily cigarette consumption, number of previous attempts to quit, CO-oximetry result in parts per million, cessation date, and level of nicotine dependence); and related variables (concomitant use of cannabis, prescribed pharmacological treatment, and type of pharmacological treatment indicated for the dishabituation process). Given the pragmatic nature of this trial, patients could contact the professionals or the chatbot at any time, and professionals could schedule follow-up visits with patients in the CG depending on both the recommendations in the portfolio of provided services and patient needs. Information regarding contact time and number of interactions was automatically recorded in the data collection notebook (CG) or by the chatbot (IG). No qualitative feedback was obtained from participants at any moment.

If the patient did not attend the 6-month follow-up visit (T1), their assigned professional tried to contact them on the phone up to 3 times to set an appointment, after which they were considered lost to follow-up.

No changes to trial outcomes were made after the trial commenced.

Sample Size

The sample size was calculated based on the outcomes of the FTFT-AP trial [22], a recent clinical trial that assessed the effectiveness of usual clinical practice in our health care system and reported a continuous abstinence rate of 9.6% in the CG at 6 months. Considering a 2-fold success rate compared with the later study [14], an α error of 5%, and a power of 80%, the calculated sample size was 418 patients. With an estimated dropout rate of 10%, the final size was 460 smokers (230 in each arm).

Statistical Analysis

Intention-to-treat analyses were performed by coding all losses to follow-up as smokers, as specified in the previously published protocol [19]. Stata 14 was used for the analyses.

Regression models (logistic or linear, as appropriate) were employed for analyzing the effect of the intervention on all outcomes and adjusted for robust estimators to account for patients recruited by clusters. Missing data were analyzed using the baseline-observation-carried-forward (BOCF) approach. The effectiveness of the intervention on the primary outcome was assessed via intergroup differences in the biochemically validated abstinence at T1 and T0, reported in proportions with the corresponding 95% CI. A sensitivity analysis was performed to compare the intention-to-treat and per-protocol analyses. Factors associated with confirmed continuous abstinence at 6 months were evaluated using a logistic regression model. Intragroup differences in quality of life between T1 and T0, as measured via the EQ-5D-5L VAS, were calculated as part of the analysis of secondary outcomes. Additionally, variables measuring the intensity of use were evaluated via intergroup differences in the average number of contacts and total interaction time. An analysis by subgroups was also conducted to account for the intensity of use with the chatbot or the contact intensity between patients and professionals.

Statistical tests for independent samples (Student *t* test and chi-square test) were applied for intergroup comparisons at baseline, and a repeated-measures ANOVA for related samples was used for evaluating intragroup differences and changes over time.

Ethics Approval

This clinical trial was approved by the Ethics Committees for Clinical Research of the Community of Madrid (December 13, 2017; approval number: 23/17) and the University Hospital 12 de Octubre (Madrid, January 30, 2018; approval number: 18/054).

Results

Characteristics of Patients

Participants were recruited between October 1, 2018, and March 31, 2019. The last follow-up visit took place on October 31, of 2019. No critical “secular events” fell into the study period. The trial ended as planned in the protocol.

A total of 161 professionals collaborated in the trial and each recruited a mean of 3.18 (SD 1.69) patients. Of the 572 potentially eligible patients who had been invited to participate, 513 accepted, provided informed consent, and were thus enrolled in the trial. Participating patients showed characteristics similar to nonparticipants in terms of gender and age.

No significant differences were found between the IG and CG at baseline in terms of sociodemographic variables or those related to their tobacco consumption (Table 1). The average age was 49.8 (SD 10.82) years, 59.3% (304/513) were women, 93.8% (481/513) were Spanish, 68.2% (350/513) had completed secondary or university education, and 51.5% (264/513) earned under €17,000/year or US $18,100/year (nearly twice the minimum wage).

Concerning variables related to tobacco use, 10.1% (52/513) of patients reported moderate or high dependence on nicotine with Heavy Smoking Index values of 4-6 points and average consumption of 16.5 cigarettes/day (SD 7.75). Additionally, 3.3% (17/513) of patients were frequent cannabis users. The mean baseline CO-oximetry level was 15.11 ppm (SD 14.12) and mean attempts to quit were 2.48 (SD 2.91). Pharmacological treatment was prescribed for 49.3% (253/513) of patients. The mean baseline score on the EQ-5D-5L VAS was slightly higher in the CG (71.8, SD 18.1), compared with that in the IG (69.4, SD 18.5; P=.07), without intergroup differences in the questionnaire dimensions expressed by their relevant weighted Health Status Index [21].

Measurements were obtained at the follow-up visit (T1) for 232 (45.2%) patients, 42.9% (104/242) and 47.2% (128/271) in the IG and CG, respectively, without significant intergroup differences (Figure 1). The analysis of dropouts also did not show significant intergroup differences (Table 2).

Table 1. Clinical characteristics of patients.

Variable	Control group (n=271)	Intervention group (n=242)	P value
Age (years), mean (SD)	50.66 (10.42)	49.01 (11.22)	.09
Gender, n (%)			.64
Women	158 (58.3)	146 (60.3)	
Men	113 (41.7)	96 (39.7)	
Educational level, n (%)			.67
Primary school or inferior	89 (32.8)	74 (30.6)	
High school	132 (48.7)	116 (47.9)	
University	50 (18.5)	52 (21.5)	
Personal gross income (€[a]/year), n (%)			.98
<8500	48 (17.7)	44 (18.2)	
8500-16,999	94 (34.7)	78 (32.2)	
17,000-25,499	73 (26.9)	66 (27.3)	
25,500-33,999	36 (13.3)	34 (14.0)	
>34,000	20 (7.4)	20 (8.3)	
Country, n (%)			.26
Spain	251 (92.6)	230 (95.0)	
Other countries	20 (7.4)	12 (5.0)	
Number of daily cigarettes, mean (SD)	16.32 (8.04)	16.70 (7.43)	.59
Previous tobacco withdrawal attempts, mean (SD)	2.37 (2.83)	2.60 (3.00)	.38
Heavy Smoking Index, mean (SD)	2.65 (1.68)	2.71 (1.59)	.67
Cannabis use, n (%)			.14
No	265 (97.8)	231 (95.5)	
Yes	6 (2.2)	11 (4.5)	
Pharmacological treatment[b], n (%)	130 (48)	119 (49.2)	.63
Simple nicotine replacement treatment	10 (7.7)	13 (10.9)	
Combined nicotine replacement treatment	4 (3.1)	5 (4.2)	
Bupropion	24 (18.5)	21 (17.6)	
Varenicline	89 (68.5)	74 (62.2)	
Others	3 (2.3)	6 (5.0)	
No pharmacological treatment	141 (52.0)	123 (50.8)	
Baseline CO-oximetry (ppt), mean (SD)	15.51 (15.33)	14.70 (12.67)	.57
EuroQol 5D-5L VAS[c], mean (SD)	71.8 (18.1)	69.4 (18.5)	.14
EuroQol 5D-5L index, mean (SD)	0.9 (0.2)	0.9 (0.2)	.11

[a]€1=US $1.06.

[b]n=130 and 119, respectively, for the IG and CG.

[c]VAS: visual analog scale.

Primary Outcome

Table 2 presents detailed intervention outcomes, from both the intention-to-treat (n=513) and per-protocol (n=232) analyses. In the intention-to-treat analysis using the BOCF at T1, an intergroup difference in the primary outcome was found, with a biochemically validated abstinence rate of 26.0% (63/242) in the IG versus 18.8% (51/271) in the CG (odds ratio [OR] 1.50, 95% CI 1.00-2.31; P=.05). After adjusting by CO-oximetry and

bupropion intake, no substantial changes were observed (OR 1.52, 95% CI 0.99-2.33; P=.053; pseudo-R^2=0.045).

In the explanative model, the factors found to correlate with the abstinence rate at 6 months were having received the chatbot intervention (OR 1.52, 95% CI 0.99-2.33; P=.053) and bupropion prescription (OR 2.81, 95% CI 1.49-5.32; P=.001). Baseline CO-oximetry level was not found to correlate with the abstinence rate at this time point (OR 0.96, 95% CI 0.94-0.99; P=.002; pseudo-R^2=0.045).

Table 2. Abstinence rate at 6 months.

Groups and rate difference	CO-validated continuous abstinence (intention to treat)	CO-validated continuous abstinence (per protocol)
Control group, n/N (%)	51/271 (18.8)	51/128 (39.8)
Intervention group, n/N (%)	63/242 (26.0)	63/104 (60.6)
Rate difference (95% CI)	–7.2 (–14.4 to 0.0)	–20.7 (–33.4 to –8.1)
Odds ratio raw (95% CI)	1.5 (1.0 to 2.3)	2.3 (1.4 to 3.9)
P value	.05	<.001
Odds ratio adjusted (95% CI)[a]	1.52 (0.99 to 2.33)	2.35 (1.37 to 4.05)
P value	.05	.002

[a]Adjusted by baseline CO-oximetry and bupropion intake.

Secondary Outcomes

In terms of quality of life, no intergroup differences were found at baseline on the VAS (71.8 in the CG versus 69.4 in the IG; P=.07). At 6 months, a significant difference on the EQ-5D-5L VAS was observed between those who had quit and those who had not (73.2 versus 64.7 points, respectively; P=.01) and also between patients in the IG and the CG (71.6 versus 66.7 points, respectively; P=.09), although statistical significance was not reached (P<.05).

In terms of variables related to intervention intensity, the mean total interaction time with the patients was 21.2 minutes (SD 18.3; 95% CI 19.0-23.4) in the CG and 121 minutes (SD 157.5; 95% CI 121.1-140.0) in the IG (P<.001), and the mean number of contacts was 2.92 (SD 1.89) in the CG and 45.56 (SD 36.32) in the IG (P<.001). Therefore, the mean interaction duration between the chatbot and patient was 2.65 minutes versus 7.26 minutes between the professional and patient. Contact was defined as the time attending consultation for cessation in the CG or as the chatbot-patient interaction plus the time for performing an activity in the IG, with a pause of more than 90 minutes being considered as the end of a contact.

Patients in the IG who had successfully quit interacted an average time of 176.1 minutes (CI 95% 124.4-227.7) versus 116.6 minutes (95% CI 65.6-167.7) for those who had not (P=.06). In the CG, the mean interaction time was 24.1 minutes (95% CI 19.1-29.2) for patients who had quit smoking versus 23.5 minutes (95% CI 19.8-27.3) for those who had not (P=.84). The average number of contacts in the IG was greater for patients who stopped smoking versus those who did not succeed (59.4 vs 40.9, respectively; P=.004), which was in contrast to the number of contacts in the CG (4.1 versus 3.6, respectively; P=.06).

An additional exploratory analysis by subgroups, which the protocol did not contemplate, was performed to assess the intensive use of the chatbot, defined as more than 4 contacts with the chatbot and over 30 minutes of total interaction time throughout the 6 months. The biochemically validated abstinence rate in the IG at T1 was significantly higher for patients who contacted the chatbot intensively versus those who did not (68.6% versus 40.9%, respectively; P=.02), which was in contrast to that observed in the CG (47.6% for patients having intensive contact with the health care worker versus 35.4% who were not; P=.30), for which also intensive contact was defined as more than 4 contacts and over 30 minutes of total interaction time throughout the 6 months.

Approximately half of the patients (130/271, 47.9% and 119/242, 49.2% in the CG and IG, respectively) received pharmacological treatment to quit smoking, with no observed intergroup differences. In the multivariate analysis, a relationship was found only between bupropion intake and biochemically confirmed abstinence at 6 months (OR 3.46, 95% CI 1.12-10.51).

Discussion

Principal Findings

Although no significant difference in smoking cessation rates was obtained, our results suggest an effect that is certainly promising (OR 1.5), with a difference in effect ranging from no effect (OR 1) or a 1% decrease (OR 0.99) in the raw result up to over 2-fold increase (OR 2.33). However, all values within the interval limits are reasonably compatible with the data, given the statistical assumptions made to calculate the interval. Therefore, these results must be interpreted with caution.

In terms of secondary variables, quality of life further improved for patients assigned to the chatbot intervention versus the CG, especially for those who succeeded in quitting. This finding is consistent with the higher abstinence rates observed in the IG and can be related to the success in quitting smoking rather than the assigned intervention. Nevertheless, the change observed in the IG, which showed a slightly lower quality of life at baseline, could be partly due to the intervention.

Both the total interaction time and the number of contacts were much greater in the IG than in the CG, although the average

contact duration was shorter in the former group. The number of sessions and invested time are key factors for the effectiveness of interventions in smokers [2]. One premise for the chatbot intervention was that its automated use would facilitate an intensive intervention of characteristics similar to face-to-face interventions but without requiring as many resources. However, the setting of a chatbot-patient interaction is very different from a visit to the doctor or nurse in terms of type of interaction, duration (the chatbot is accessed easily and the intervention can be fragmented according to the patient needs), and activities performed by the patient (patients in the IG practiced behavioral techniques during the intervention time, whereas those in the CG did it at home and the invested time was not registered). This could partly justify the paradox that patients in the IG who did not succeed in quitting smoking spent more time contacting the chatbot than those who did quit in the CG, although further trials are required to clarify this aspect.

Patients in the IG who made intensive use of the chatbot (longer total interaction time and greater number of contacts) achieved significantly higher abstinence rates than those who did not, contrary to the CG, where no significant differences in abstinence rates were found between those having and not having intensive interaction. It appears that the majority of professionals conducted very homogeneous interventions in the CG, probably limited by their workload. However, the number of patients in the CG achieving continuous abstinence at 6 months was higher when the interventions reached intensive use (47 versus 35, respectively), despite not reaching clinical significance, probably due to the limited sample size.

The use of pharmacological treatment for tobacco cessation in usual practice yields over a 2-fold success rate for the same intervention duration [2]. In this trial, first-choice drugs were equally prescribed in both arms and the performed analysis showed that their effect was not considerable. In any event, the chatbot was designed to increase adherence to medication by providing accessible and tailored information, although this could not be properly measured due to the low number of follow-up visits in the CG.

The per-protocol analysis revealed a difference compared with the intention-to-treat analysis (Table 2), supporting the effectiveness of Dejal@bot, yet raising concerns about adherence to the chatbot, an aspect that must be improved.

In summary, accessibility, simplicity, ubiquity, and immediacy were components that probably favored longer interaction time between the chatbot and patients and a higher number of contacts, which are key factors for predicting long-term abstinence in interventions in smokers [2,3]. These, in addition to following usual practice guidelines, were the factors underlying the effectiveness of Dejal@bot.

Further trials are required to determine the components that mainly impact the effectiveness of the chatbot and which type of patients are susceptible to benefit from this type of intervention. Besides, more studies are needed with direct technical assistance for improving accessibility, as well as interventions for improving digital competencies in certain population groups, which would likely improve retention.

Comparison With Other Studies

We identified only 2 clinical trials [16,23] that used chatbots to help people quit smoking. One study [16], without published outcomes at the time of this writing, aims at comparing a CG using SMS text messages with an IG using a chatbot (QuitBot) in smartphones for helping with the cessation process. Abstinence will be checked at 3, 6, and 12 months after the intervention via biochemical validation. The other study [23] was an experimental trial that added a chatbot to the already existing “Smoke Free App.” The trial compared the interaction between the user and the app with and without the chatbot. The inclusion of the chatbot in the app increased the self-reported abstinence at 1 month (OR 1.36, 95% CI 1.16-1.61; P<.001). Therefore, Dejal@bot is the first published clinical trial about the effectiveness of a chatbot for helping smokers to quit with biochemically validated abstinence outcomes in the long term. Multimedia Appendix 7 presents a list of articles on the use of apps and chatbots in the tobacco cessation process.

Given the absence of further similar studies, the outcomes of this trial were compared against those in several clinical trials using apps for helping to quit smoking. One study compared the effectiveness of 2 apps [24], one of them following the US clinical practice guideline [2] that achieved a 21.1% abstinence rate at 12 months versus an Acceptance and Commitment Therapy app that achieved a higher abstinence rate of 28.2% (OR 1.49, 95% CI 1.22-1.83; P<.001). These were self-reported and 1-time outcomes, unlike those in our trial.

In the study by Pallejà-Millán et al [25], participants in the IG who used the app regularly and correctly had a higher probability of being abstinent at 12 months (OR 7.20, 95% CI 2.14-24.20; P=.001) than those in the CG. That is the only trial comparing the use of an mHealth intervention with usual practice but, unlike ours, their design was based on conglomerates (health care centers) and not pragmatic. The obtained abstinence outcomes were statistically significant when contrasting correct versus incorrect use of the app, but not in the intergroup comparison. Of note, 34.2% (97/284) of patients in the IG did not enter the app for smoking cessation.

BinDhim et al [26] compared an interactive app (including a variety of options for cessation, evaluation of risks and benefits from quitting, motivational messaging, and diary of the cessation process) with a merely informative app. The abstinence results at 6 months were better with the intervention app (10.2% vs 4.8%; RR 2.02, 95% CI 1.08-3.81), although abstinence was self-reported.

Baskerville et al [27] compared an evidence-informed app for smoking cessation with an evidence-informed self-help guide for reducing the smoking prevalence among young adult smokers, and observed no differences at 6 months (OR 0.83, 95% CI 0.59-1.18). Of note, the follow-up rate was 60.48% (967/1599) at that time point.

Strengths and Limitations

Among the strengths of this study are its pragmatic design, with real-life conditions of clinical practice in terms of recruitment (inclusion criteria for patients and professionals), prescribed medication (patients were treated by their assigned practitioners

at their usual consultations, without further restrictions), and minimum number of mandatory visits (baseline and at 6 months). Computer or internet literacy of patients was not checked at any point, and randomization was only conducted after their inclusion in the trial and collection of baseline data. The fact that professionals volunteered to participate can be detrimental to the outcomes because of a possible self-selection bias (participating workers may not be representative of the health staff due to a greater interest in tobacco addiction).

Usual practice was selected as the comparator due to being the standard treatment in our setting and because the chatbot attempts to reproduce the ideal face-to-face interaction between therapists and patients but as a novel technological support. This led to comparing 2 interventions of different intensities in terms of interaction time and number of contacts between the health worker or the chatbot and the patient. However, this comparison was justified by the pragmatic nature of the study.

The main outcome variable (continuous abstinence at 6 months) was biochemically validated, which increased the scientific accuracy of the results. So far, all clinical trials with apps [24,26,27] or chatbots [23] considered a patient to be abstinent based only on self-reports, with the exception of Pallejà-Millán et al [25].

The mentioned strengths reinforce the validity of the external outcomes, especially for health systems similar to the Spanish public health service. Although the Dejal@bot intervention cannot be directly delivered to the internet community without the intervention of a health professional to prescribe medication (if indicated), it could be provided by public or private health insurance systems. Alternatively, the pharmacological component of the intervention can be omitted to be able to implement it without the need for a health professional.

In terms of applicability, the system is ready for use and has enormous potential scalability, which could be improved with personalized technical assistance to facilitate accessibility, a key factor affecting the outcomes.

The main limitation of this trial was the dropout rate of 54.8% (281/513). Given the pragmatic design of the trial, no further midterm reinforcements or visits could be scheduled. Additionally, 38.8% (94/242) of the IG users never entered the chatbot. Losses to follow-up were homogeneous in both study arms, both quantitatively and in terms of participants' characteristics after the intervention.

Implications of the Study Findings/Implications of All Available Evidence

Dejal@bot showed its effectiveness in increasing nicotine abstinence rates in the long term compared with standard treatment provided by the usual doctor or nurse assigned to the patient, although these results must be interpreted with caution given the high dropout rate.

This intervention can facilitate patient access to high-quality treatments for the leading cause of preventable death (ie, tobacco smoking), saving costs for the health provider and reducing the workload for the professionals. At the time of this writing, with reduced mobility and social distancing due to the COVID-19 pandemic, this was especially appropriate and pertinent and could make a difference in the population's health.

Further evidence is still required to assess the effectiveness of mHealth in smoking cessation. Although there are trials assessing the use of SMS text messages and apps for quitting, interventions using chatbots need to be evaluated, and qualitative studies about cost-effectiveness, usability, and satisfaction must be conducted. Additionally, determining the components that mainly affect effectiveness will be of interest to achieve behavioral changes and increased participation of users, because a strong association appears to exist between the time of use or accomplishment of tasks and dropout rates.

From the ethics perspective, the importance of high-quality studies evaluating these treatments must be highlighted, which will prevent the patient from being disfavored by incomplete, biased, or nonevidence-based interventions, and will also avoid decreased accessibility of certain population segments to quality therapies.

Acknowledgments

We are grateful for the contributions of the late Esperanza Escortell-Mayor to the study's design. We miss you.

This study has been funded by Instituto de Salud Carlos III through the project "PI17/01942", as part of the Plan Estatal de I+D+I 2013-2016 co-funded by European Regional Development Fund (ERDF) "A way of shaping Europe." The main researcher (EOE) received a grant from the Fundación para la Investigación e Innovación Biomédica en Atención Primaria (FIIBAP) through their 2018 call for grants for promoting research programs. The funders had no role in the development of the study. The Dejal@ Group (collaborative author) included the following collaborators: Molina Alameda L; Andrade Rosa C; Espejo Romero M; Hernandez Garcia JL; Perez Perez E; Zambrano Alvarez J; Sanchez Jimenez FJ; Garcia Silva M; Martin Gomez I; Puyo Rodriguez N; Sanchez Vazquez I; Garcia Alegre MT; Moral Moraleda C; Perez Gandia O; Hernandez Rodriguez S; Aparicio Velasco J; Cerezo Pascual E; Marin Osorio C; Prieto Villegas E; Rodriguez Donoso J; Rodriguez Huete A; Romero Blanco FJ; Castillo Portales SJ; Collados Navas R; Gutierrez Sordo P; Sanchez Quejido M; Abad Schilling C; Capitan Jurado M; Regalado del Valle MA; Alache Zuñiga HC; Jaenes Barrios B; Ortiz Sánchez M; Santiago Hernando ML; Arjona Pérez S; Díaz Martin R; Fandiño García B; Gómez Ortiz MC; López-Martin Aragón MT; Miguel Abanto MA; Noguerol Álvarez M; Parra Roman S; Rey Rodríguez S; Aguilar Hurtado E; Aguilera Rubio M; Carbonero Martín AI; Esteban García M; Esteban De la Peña C; García Contreras C; Rojas Giraldo MJ; Sánchez Fernández YB; Alonso Ovies M; García Arpa N; González-Carpio Paredes O; Grandes Muñoz MJ; Jara Peñacoba M; Lopez Villalvilla A; Parellada Ruiz de la Cuesta MJ; Rodríguez Bastida M; Vicente Pérez S; Martín Porras B;

González Valls A; Joglar Alcubilla V; Santa Cruz Hernández J; Castilla Álvarez C; García Viada M; García-Abad Fernández MA; Molina Bárcena V; Sepúlveda Gómez I; Cuadrado González P; Linares Sánchez C; Solans Aisa B; Francisco Romanillos T; García Ortego A; Hernández García-Alcalá P; Huguet Vivas F; Juarez Zapatero ML; López García C; Manzano Martín MJ; Martínez Torres JA; Miguel Garzón M; Rodrigo Rodrigo MP; Ruano Domínguez MA; Aguilera Lafuente D; Altares Arriola N; Aranzo Pacheco R; Escudero Araus M; Fuentes Manrique C; Guijarro Abanades S; López Carabaño AM; Moneva Vicente GD; Pérez Fernández M; Rubio Rubio T; Venegas Gato MF; Blas Escribano M; Castro Fouz MM; Castro Sánchez B; Domínguez Pérez L; García Gómez PA; Herrero Dios A; Lor Leandro M; Lozano Marín I; Pacho Pinto S; Pastor Sánchez R; Moreno Chaparro MD; Yagüe Fernández E; Mateo Madruga A; Pumar Sainz P; Aragon Marente C; Gomez Medina MA; Granados Garrido JA; Gutierrez Sanchez I; Heras Alonso MJ; Herrero Pacheco MR; Martín Peces B; Misericordia Garcia S; Ochoa Vilor S; Vargas Machuca C; González Bellido MA; Goncalves Molina B; Morata Cruz F; Moreno Carreras L; Alda de la Fuente MV; Hebeisen Andrés L; Fernández-Montes López-Morato O; Marcos de Frutos C; Ruiz Pascual V; Cayuela Maté A; Fernández García RM; Góngora Marín A; Perez Cuadrado S; Redondo Horcajo A; Rosario Pacheco IÁ; Villena Romero RM; Hakami Hakami O; Álvarez Villalba MM; Bedoya Frutos MJ; Hormigos Agraz A; López López MT; Perez Vallejo B; Tardaguila Lobato MP; Ballarín González A; Fernández Fernández Y; Ferrer Zapata I; Gómez Suárez E; Morales Ortiz F; Peláez Laguno C; Quintana Gómez JL; Rioja Delgado E; Aza Pascual JI; Fuente Arriarán D; García Pérez RA; Gárriz Aguirre A; Gómez Díaz S; Gonzalez Sanchez N; Sanz Velasco C; Vázquez Gallego J; Fernandez Durán C; Peña Antón N; Blanco Canseco JM; Minguela Puras E.

Authors' Contributions

JFA-T, EO-E, and CM-L conceived the study and participated in its design and coordination. FJM-S, BM-P, and MES-S participated in different phases of the protocol design and coordination. Clinical researchers in clinical practice of the Dejal@bot GROUP conducted the fieldwork in their health centers. IDC-G and MG-C performed the statistical analyses. The first draft was initially written by JFA-T, EO-E, and CM-L, which was then discussed and rewritten together with IDC-G. FJM-S, BM-P, and MES-S collaborated in the writing of the manuscript. All authors contributed to the data interpretation, critically reviewed the first draft, approved the final version, and agreed to be accountable for the work.

Conflicts of Interest

JFA-T, EO-E, and CM-L designed the chatbot and own the intellectual property rights. The remaining research team members (authors, developers, and sponsors) were not involved in the development of the intervention.

Multimedia Appendix 1
Differences between an app and a chatbot.
[PNG File , 713 KB - mhealth_v10i6e34273_app1.png]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 411 KB - mhealth_v10i6e34273_app2.pdf]

Multimedia Appendix 3
Documento de información para el paciente.
[DOCX File , 32 KB - mhealth_v10i6e34273_app3.docx]

Multimedia Appendix 4
Documento de Consentimiento Informado.
[OCX File , 26 KB - mhealth_v10i6e34273_fig.ocx]

Multimedia Appendix 5
Screen capture.
[PNG File , 740 KB - mhealth_v10i6e34273_app5.png]

Multimedia Appendix 6
TIDieR checklist.
[DOCX File , 30 KB - mhealth_v10i6e34273_app6.docx]

Multimedia Appendix 7
Published articles about apps and chatbots to help in the tobacco cessation process.
[PDF File (Adobe PDF File), 359 KB - mhealth_v10i6e34273_app7.pdf]

References

1. World Health Organization. Tobacco: Key Facts. World Health Organization. 2019 Mar 9. URL: https://www.who.int/news-room/fact-sheets/detail/tobacco [accessed 2022-05-28]
2. Clinical Practice Guideline Treating Tobacco UseDependence 2008 Update Panel, Liaisons,Staff. A clinical practice guideline for treating tobacco use and dependence: 2008 update. A U.S. Public Health Service report. Am J Prev Med 2008 Aug;35(2):158-176 [FREE Full text] [doi: 10.1016/j.amepre.2008.04.009] [Medline: 18617085]
3. Aveyard P, West R. Managing smoking cessation. BMJ 2007 Jul 07;335(7609):37-41 [FREE Full text] [doi: 10.1136/bmj.39252.591806.47] [Medline: 17615224]
4. Olano-Espinosa E. Efectividad de la formación sobre abordaje del tabaquismo en los profesionales sanitarios de Atención Primaria de Salud: un ensayo clínico aleatorizado y controlado [dissertation]. Departamento de Medicina Interna, Universidad Complutense de Madrid. 2011. URL: https://eprints.ucm.es/id/eprint/14452/1/T33348.pdf [accessed 2021-06-06]
5. Bartsch AL, Härter M, Niedrich J, Brütt AL, Buchholz A. A Systematic Literature Review of Self-Reported Smoking Cessation Counseling by Primary Care Physicians. PLoS One 2016;11(12):e0168482 [FREE Full text] [doi: 10.1371/journal.pone.0168482] [Medline: 28002498]
6. Hedman L, Katsaounou PA, Filippidis F, Ravara S, Lindberg A, Janson C, EUREST-PLUS consortium. Receiving support to quit smoking and quit attempts among smokers with and without smoking related diseases: Findings from the EUREST-PLUS ITC Europe Surveys. Tob Induc Dis 2018;16:A14 [FREE Full text] [doi: 10.18332/tid/102787] [Medline: 31516468]
7. Saito A, Nishina M, Murai K, Mizuno A, Ueshima F, Makiishi T, et al. Health professional's perceptions of and potential barriers to smoking cessation care: a survey study at a dental school hospital in Japan. BMC Res Notes 2010 Dec 07;3(1):329 [FREE Full text] [doi: 10.1186/1756-0500-3-329] [Medline: 21138553]
8. Taylor G, Dalili M, Semwal M, Civljak M, Sheikh A, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2017 Sep 04;9:CD007078 [FREE Full text] [doi: 10.1002/14651858.CD007078.pub5] [Medline: 28869775]
9. WHO Global Observatory for eHealth. MHealth: new horizons for health through mobile technologies. WHO. Geneva, Switzerland: World Health Organization; 2011. URL: http://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2022-05-31]
10. Global System for Mobile Communications Association (GSMA). The Mobile Economy Report. GSMA. URL: https://www.gsma.com/mobileeconomy/wp-content/uploads/2020/03/GSMA_MobileEconomy2020_Global.pdf [accessed 2021-05-29]
11. Abroms LC, Lee Westmaas J, Bontemps-Jones J, Ramani R, Mellerson J. A content analysis of popular smartphone apps for smoking cessation. Am J Prev Med 2013 Dec;45(6):732-736 [FREE Full text] [doi: 10.1016/j.amepre.2013.07.008] [Medline: 24237915]
12. Verheijden MW, Jans MP, Hildebrandt VH, Hopman-Rock M. Rates and determinants of repeated participation in a web-based behavior change program for healthy body weight and healthy lifestyle. J Med Internet Res 2007 Jan 22;9(1):e1 [FREE Full text] [doi: 10.2196/jmir.9.1.e1] [Medline: 17478410]
13. National Institute for Health and Care Excellence (NICE). NICE Guideline [NG183]: Behaviour Change: Digital and Mobile Health Interventions. National Institute for Health and Care Excellence (NICE). 2020 Oct. URL: https://www.nice.org.uk/guidance/ng183 [accessed 2022-05-28]
14. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y, Dobson R. Mobile phone text messaging and app-based interventions for smoking cessation. Cochrane Database Syst Rev 2019 Oct 22;10:CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub5] [Medline: 31638271]
15. Cobos-Campos R, de Lafuente AS, Apiñaniz A, Parraza N, Llanos IP, Orive G. Effectiveness of mobile applications to quit smoking: Systematic review and meta-analysis. Tob Prev Cessat 2020;6:62 [FREE Full text] [doi: 10.18332/tpc/127770] [Medline: 33241162]
16. Fred Hutchinson Cancer Research Center. Full Scale Randomized Trial of an Innovative Conversational Agent for Smoking Cessation (QuitBot). ClinicalTrials.gov. 2020. URL: https://clinicaltrials.gov/ct2/show/NCT04308759 [accessed 2021-01-06]
17. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340(mar23 1):c869-c869 [FREE Full text] [doi: 10.1136/bmj.c869] [Medline: 20332511]
18. Turner L, Shamseer L, Altman D, Weeks L, Peters J, Kober T, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. Cochrane Database Syst Rev 2012 Nov 14;11:MR000030 [FREE Full text] [doi: 10.1002/14651858.MR000030.pub2] [Medline: 23152285]
19. Avila-Tomas JF, Olano-Espinosa E, Minué-Lorenzo C, Martinez-Suberbiola FJ, Matilla-Pardo B, Serrano-Serrano ME, Group Dej@lo. Effectiveness of a chat-bot for the adult population to quit smoking: protocol of a pragmatic clinical trial in primary care (Dejal@). BMC Med Inform Decis Mak 2019 Dec 03;19(1):249 [FREE Full text] [doi: 10.1186/s12911-019-0972-z] [Medline: 31796061]
20. West RHP, Hajek P, Stead L, Stapleton J. Outcome criteria in smoking cessation trials: proposal for a common standard. Addiction 2005 Mar;100(3):299-303. [doi: 10.1111/j.1360-0443.2004.00995.x] [Medline: 15733243]

21. Stolk E, Ludwig K, Rand K, van Hout B, Ramos-Goñi JM. Overview, Update, and Lessons Learned From the International EQ-5D-5L Valuation Work: Version 2 of the EQ-5D-5L Valuation Protocol. Value Health 2019 Jan;22(1):23-30 [FREE Full text] [doi: 10.1016/j.jval.2018.05.010] [Medline: 30661630]
22. Minué-Lorenzo C, Olano-Espinosa E, Del Cura-González I, Vizcaíno-Sánchez JM, Camarelles-Guillem F, Granados-Garrido J, et al. Subsidized pharmacological treatment for smoking cessation by the Spanish public health system: A randomized, pragmatic, clinical trial by clusters. Tob Induc Dis 2019 Sep 5;17(September):64 [FREE Full text] [doi: 10.18332/tid/111368] [Medline: 31582953]
23. Perski O, Crane D, Beard E, Brown J. Does the addition of a supportive chatbot promote user engagement with a smoking cessation app? An experimental study. Digital Health 2019 Sep 30;5:205520761988067. [doi: 10.1177/2055207619880676]
24. Bricker JB, Watson NL, Mull KE, Sullivan BM, Heffner JL. Efficacy of Smartphone Applications for Smoking Cessation: A Randomized Clinical Trial. JAMA Intern Med 2020 Nov 01;180(11):1472-1480 [FREE Full text] [doi: 10.1001/jamainternmed.2020.4055] [Medline: 32955554]
25. Pallejà-Millán M, Rey-Reñones C, Barrera Uriarte ML, Granado-Font E, Basora J, Flores-Mateo G, et al. Evaluation of the Tobbstop Mobile App for Smoking Cessation: Cluster Randomized Controlled Clinical Trial. JMIR Mhealth Uhealth 2020 Jun 26;8(6):e15951 [FREE Full text] [doi: 10.2196/15951] [Medline: 32589153]
26. BinDhim NF, McGeechan K, Trevena L. Assessing the effect of an interactive decision-aid smartphone smoking cessation application (app) on quit rates: a double-blind automated randomised control trial protocol. BMJ Open 2014 Jul 18;4(7):e005371 [FREE Full text] [doi: 10.1136/bmjopen-2014-005371] [Medline: 25037644]
27. Baskerville NB, Struik LL, Guindon GE, Norman CD, Whittaker R, Burns C, et al. Effect of a Mobile Phone Intervention on Quitting Smoking in a Young Adult Population of Smokers: Randomized Controlled Trial. JMIR Mhealth Uhealth 2018 Oct 23;6(10):e10893 [FREE Full text] [doi: 10.2196/10893] [Medline: 30355563]

Abbreviations

AI: artificial intelligence
BOCF: baseline-observation-carried-forward
CG: control group
CONSORT: Consolidated Standards of Reporting Trials
ICT: information and communications technology
IG: intervention group
mHealth: mobile health
OR: odds ratio
ppm: parts per million
RR: relative risk
TIDieR: Template for Intervention Description and Replication
VAS: visual analog scale

Edited by L Buis; submitted 18.10.21; peer-reviewed by J Bricker, J Ropero; comments to author 27.12.21; revised version received 20.02.22; accepted 03.05.22; published 27.06.22.

Please cite as:
Olano-Espinosa E, Avila-Tomas JF, Minue-Lorenzo C, Matilla-Pardo B, Serrano Serrano ME, Martinez-Suberviola FJ, Gil-Conesa M, Del Cura-González I, Dejal@ Group
Effectiveness of a Conversational Chatbot (Dejal@bot) for the Adult Population to Quit Smoking: Pragmatic, Multicenter, Controlled, Randomized Clinical Trial in Primary Care
JMIR Mhealth Uhealth 2022;10(6):e34273
URL: https://mhealth.jmir.org/2022/6/e34273
doi:10.2196/34273
PMID:35759328

Original Paper

Assessing the Risk Factors For Diagnosed Symptomatic Dry Eye Using a Smartphone App: Cross-sectional Study

Ngamjit Kasetsuwan[1,2,3], MD; Olan Suwan-Apichon[2,4], MD; Kaevalin Lekhanont[2,5], MD; Varintorn Chuckpaiwong[2,5], MD; Usanee Reinprayoon[1,2,3], MD; Somporn Chantra[2,6], MD; Vilavun Puangsricharern[1,2,3], MD; Lalida Pariyakanok[2,7], MD; Pinnita Prabhasawat[2,8], MD; Nattaporn Tesavibul[2,8], MD; Winai Chaidaroon[2,9], MD; Napaporn Tananuvat[2,9], MD; Chakree Hirunpat[2,10], MD; Nauljira Prakairungthong[2,11], MD; Wiwan Sansanayudh[2,12], MD; Chareenun Chirapapaisan[2,8], MD; Pakornkit Phrueksaudomchai[2,13], MD

[1]Department of Ophthalmology, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand
[2]Cornea and Refractive Surgery Society of Thailand, Department of Ophthalmology, King Chulalongkorn Memorial Hospital, Bangkok, Thailand
[3]Excellence Center of Cornea and Limbal Stem Cell Transplantation, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand
[4]Department of Ophthalmology, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand
[5]Department of Ophthalmology, Faculty of Medicine, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand
[6]Department of Ophthalmology, Rajavithi Hospital, Bangkok, Thailand
[7]Thai Red Cross Eye Bank, Bangkok, Thailand
[8]Department of Ophthalmology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand
[9]Department of Ophthalmology, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand
[10]Department of Ophthalmology, Faculty of Medicine, Songklanagarind Hospital, Prince of Songkla University, Songkhla, Thailand
[11]Department of Ophthalmology, Mettapracharak Hospital, Nakhon Pathom, Thailand
[12]Department of Ophthalmology, Phramongkutklao College of Medicine, Bangkok, Thailand
[13]Department of Ophthalmology, Faculty of Medicine, Thammasat University, Pathum Thani, Thailand

Corresponding Author:
Ngamjit Kasetsuwan, MD
Department of Ophthalmology
Faculty of Medicine
Chulalongkorn University
Rama 4 Rd Pathum Wan
Pathumwan District
Bangkok, 10330
Thailand
Phone: 66 22564000 ext 81875
Email: ngamjitk@gmail.com

Abstract

Background: Dry eye (DE) is a chronic inflammatory disease of the ocular surface of the eye that affects millions of people throughout the world. Smartphone use as an effective health care tool has grown exponentially. The “Dry eye or not?” app was created to evaluate the prevalence of symptomatic DE, screen for its occurrence, and provide feedback to users with symptomatic DE throughout Thailand.

Objective: The purpose of this study was to compare the prevalence of symptomatic dry eye (DE), blink rate, maximum blink interval (MBI), and best spectacle-corrected visual acuity (BSCVA) between people with and without symptomatic DE and to identify risk factors for symptomatic DE in Thailand.

Methods: This cross-sectional study sourced data from the “Dry eye or not?” smartphone app between November 2019 and July 2020. This app collected demographic data, Ocular Surface Disease Index (OSDI) score, blink rate, MBI, BSCVA, and visual display terminal (VDT) use data. The criterion for symptomatic DE was OSDI score ≥13.

Results: The prevalence of symptomatic DE among individuals using this smartphone app in Thailand was 85.8% (8131/9482), with the Northeastern region of Thailand having the highest prevalence, followed by the Northern region. Worse BSCVA (median 0.20, IQR 0.40; P=.02), increased blink rate (median 18, IQR 16; P<.001), reduced MBI (median 8.90, IQR 10.80; P<.001),

female sex (adjusted OR 1.83; 95% CI 1.59-2.09; *P*<.001), more than 6 hours of VDT use (adjusted OR 1.59; 95% CI 1.15-2.19; *P*=.004), and lower than bachelor's degree (adjusted OR 1.30; 95% CI 1.03-1.64; *P*=.02) were significantly associated with symptomatic DE. An age over 50 years (adjusted OR 0.77; 95% CI 0.60-0.99) was significantly less associated with symptomatic DE (*P*=.04).

Conclusions: This smartphone DE app showed that the prevalence of symptomatic DE in Thailand was 85.8%. Signs and risk factors could be also evaluated with this smartphone DE app. Screening for DE by this app may allow for the development of strategic plans for health care systems in Thailand.

(JMIR Mhealth Uhealth 2022;10(6):e31011) doi:10.2196/31011

KEYWORDS

blink rate; dry eye; smartphone application; maximum blink interval; prevalence; mHealth; epidemiology; screening; risk factors; symptoms; ophthalmology; vision

Introduction

Dry eye (DE) is a chronic inflammatory disease with multifactorial etiology involving a loss of tear film homeostasis, leading to tear film instability and hyperosmolarity and triggering a vicious cycle of DE [1]. This condition affects millions of people globally with a prevalence in the range of 5%-30% in those over 50 years of age [2]. Previous research has identified many risk factors for DE, including older age, female sex, refractive surgery, connective tissue disease, low humidity environment, and use of visual display terminals (VDTs) [2]. Symptoms of dry eye are varied and include itching, burning, stinging, pain, photophobia, foreign body sensation, ocular redness, and blurred vision. Despite this, many people with DE symptoms remain unevaluated, undiagnosed, and untreated [3].

Smartphone use has grown globally at an exponential rate and has been proven as an effective health care tool for use by patients and physicians [4]. Many smartphone apps have been developed to support and empower patients, including apps for DE screening that evaluate lifestyle and associated risk factors [5-13]. We designed the "Dry eye or not?" app using Flutter by Google to identify individuals with a diagnosis of symptomatic DE, document DE symptoms, and assess blink rate, maximum blink interval (MBI) [14], best spectacle-corrected visual acuity (BSCVA), and risk factors associated with diagnosed symptomatic DE. This app was created by the Cornea and Refractive Surgery Society of Thailand. Because it was easy access and people in any region of the country could download this app, it was a convenient tool for evaluating the prevalence of symptomatic DE throughout the country and for screening and providing feedback to users with symptomatic DE, such as clinical advice.

The aim of this study was to estimate symptomatic DE prevalence and compare prevalence among regions of Thailand. In addition, this study aimed to compare blink rate, MBI, and BSCVA between individuals with and without symptomatic DE and to identify risk factors for this condition using "Dry eye or not?" app.

Methods

Study Participants

This cross-sectional study used the custom-designed "Dry eye or not?" smartphone app that was available for download in Thailand from November 2019 to July 2020. The app was released free of charge by the Cornea and Refractive Surgery Society of Thailand, a group of cornea and ocular surface disease experts, with no financial compensation. All voluntary users gave informed consent in electronic format. The inclusion criteria included individuals who were be able to read Thai language and had smartphones. Incomplete responses for blink rate, maximum blink interval, BSCVA, and the OSDI questionnaire were excluded.

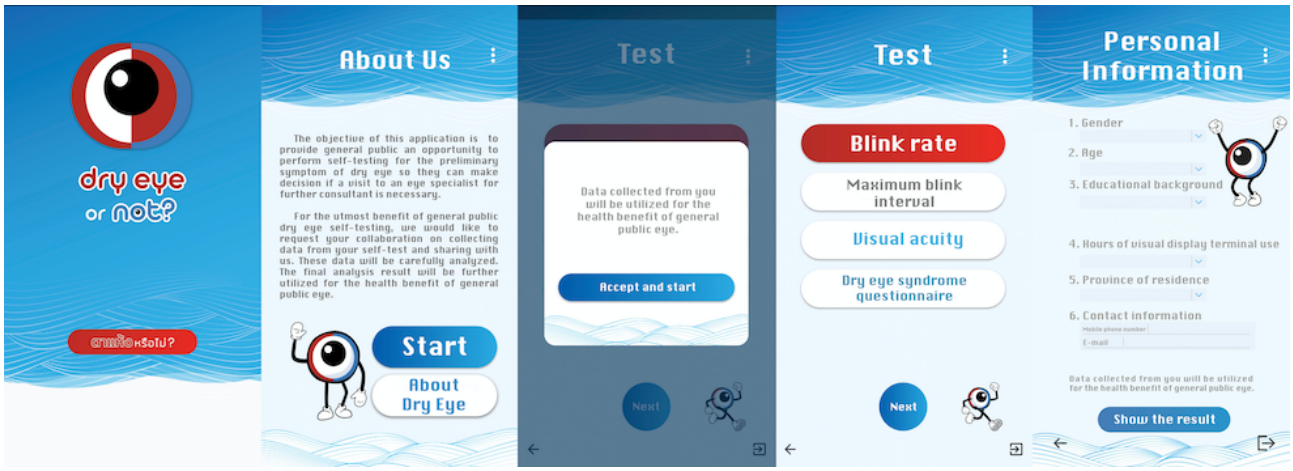
Ethics Approval

This study followed the tenets of the Declaration of Helsinki and was approved by the Institutional Review Board of the Faculty of Medicine, Chulalongkorn University, Thailand.

Data Collection

The app was freely downloadable via smartphone-based iOS and Android operating systems. This app collected data on blink rate (per minute), maximum blink interval [14] (seconds, secs), BSCVA (logMAR), DE symptoms, and demographic characteristics, as shown in Figure 1. The application programming interface (API) of blink detection in this app was used with a machine learning (ML) face detection kit developed by Google, which can recognize, locate, and determine the contours of facial features. We used this API to detect and record the number of eye blinks and MBI. Test instructions were displayed on smartphones before each test started, and the front camera was automatically accessed. Users were instructed to fit their face image to the camera display with a viewing distance of about 40 cm. Symptoms of DE were evaluated using the 12-item Ocular Surface Disease Index (OSDI) (Multimedia Appendix 1), with scores of ≥13 diagnostic of DE and severity classified as mild (13-22 points), moderate (23-32 points), and severe (33-100 points) [15]. Demographic data included age, sex, educational level (relative to bachelor's degree), hours of VDT use per day, and region where each user was living.

Figure 1. Screenshots of "Dry eye or not" app. Left to right: welcome screen, screen of information about this app, inform consent screen, measurement screen (blink rate, maximum blink interval, best spectacle-corrected visual acuity [BSCVA], Ocular Surface Disease Index [OSDI] questionnaire), and demographic characteristics of participants.



Blink Rate and Maximum Blink Interval

Before commencing these tests, users were instructed to remove any spectacles and to blink normally and naturally. The blink test recorded the number of blinks per 30-second period and stored this as the number of blinks per minute. Before starting the MBI test, users were instructed to close their eyes in preparation and then to open them and keep them open for as long as possible. The time recording began when the eyes were opened, and the recorded duration was from this point to the first eye closure [14].

Best Spectacle-Corrected Visual Acuity

All participants were asked to wear their glasses before starting the test. The test began with tumbling E at a size equivalent to 20/40 Snellen. If the participant failed to select the correct answer, chose to skip the question, or did not provide any answer within 2 seconds, the tumbling E size was increased by 1 Snellen line. However, if the participant answered correctly on 2 consecutive occasions, the tumbling E became smaller by 1 Snellen line.

Statistical Analysis

Demographics and baseline clinical characteristics were analyzed in frequency and percentage. The categorical data were compared using Pearson chi-square test. Blink rate, MBI, and BSCVA (logMAR) were compared between individuals with and without symptomatic DE using a Wilcoxon rank-sum test and presented as median with interquartile range (IQR). Moreover, an unpaired Student *t* test was also used to compared OSDI score between groups and presented as mean with standard deviation. Kruskal-Wallis with Dunn test for multiple comparison was used to compare the BSCVA between individuals with different severity levels of symptomatic DE and without symptomatic DE. Univariate and multivariate logistic regressions, presented in crude and adjusted odds ratio (OR) with 95% CI, were used to assess the associations between risk factors and symptomatic DE. For all analyses, a *P* value of .05 was the criterion for statistical significance, and Stata software version 15.1 (StataCorp) was used.

Results

Initial Findings

A total of 13,228 individuals used the app. All were volunteers aged above 15 years of age and were Thai citizens living in Thailand. However, data from 3746 users were incomplete and were thus excluded from analysis. The complete data of the excluded participants included demographic and baseline characteristics and OSDI scores. A sensitivity analysis of complete data between included and excluded participants was done and is shown in Multimedia Appendix 2. There was a statistically significant difference in age factor ($P<.001$) and OSDI scores ($P<.001$). However, the percentage of participants in each age group was similarly distributed, and the OSDI scores between the included group (mean 30.59, SD 17.94) and the excluded group (mean 33.68, SD 19.15) was not clinically significantly different. As a result, data from 9482 users were analyzed. Of these, 1811 (19.1%) were men and 7671 (80.9%) were women. The baseline characteristics of the participants, including age, VDT use per day, educational level, and regions of residence including Northern, Northeastern, Eastern, Western, Central, Southern, and capital city (Bangkok), are shown in Table 1. A comparison of these characteristics between individuals with and without symptomatic DE is presented in Table 2.

Table 1. Demographic and baseline characteristics of participants (N=9482).

Characteristics	Values, n (%)
Age (years)	
15-20	1883 (19.9)
21-30	4612 (48.6)
31-40	1342 (14.2)
41-50	803 (8.5)
>50	842 (8.9)
VDT[a] use per day	
Less than 1 hour	278 (2.9)
1-4 hours	985 (10.4)
>4-6 hours	2432 (25.7)
>6-8 hours	2754 (29)
>8 hours	3033 (32)
Educational level	
Lower than bachelor's degree	1831 (19.3)
Bachelor's degree	6281 (66.3)
Higher than bachelor's degree	1370 (14.5)
Regions of participants' residence	
Northern region	727 (7.8)
Northeastern region	934 (10)
Central region excluding Bangkok city	1,975 (21.1)
Bangkok city	4316 (46.1)
Eastern region	656 (7)
Western region	196 (2.1)
Southern region	561 (6)

[a]VDT: visual display terminal.

Table 2. Comparison of characteristics of participants with and without symptomatic dry eye.

Characteristics	Symptomatic dry eye (n, %)		*P* value
	Without	With	
Age (years)			<.001
15-20	231 (17.1)	1652 (20.3)	
21-30	566 (41.9)	4046 (49.8)	
31-40	230 (17)	1112 (13.7)	
41-50	134 (9.9)	669 (8.2)	
>50	190 (14.1)	652 (8)	
Sex			<.001
Female	945 (69.9)	6726 (82.7)	
Male	406 (30.1)	1405 (17.3)	
VDT[a] use (hours/day)			<.001
Less than 1 hour	61 (4.5)	217 (2.7)	
1-4 hours	199 (14.7)	786 (9.7)	
>4-6 hours	392 (29)	2040 (25.1)	
>6-8 hours	360 (26.7)	2694 (29.4)	
>8 hours	339 (25.1)	2694 (33.1)	
Educational level			<.001
Lower than bachelor's degree	237 (17.5)	1594 (19.6)	
Bachelor's degree	851 (63)	5430 (66.8)	
Higher than bachelor's degree	263 (19.5)	1107 (13.6)	

[a]VDT: visual display terminal.

Prevalence of Symptomatic Dry Eye

Of the 9482 participants, 8131 (85.8%) were diagnosed with symptomatic DE. The prevalence differed significantly between regions (*P*<.001), as shown in Table 3. In addition, the prevalence of subgroups of symptomatic DE (normal, mild, moderate, and severe grade) were also significantly different among regions (*P*<.001), as shown in Table 4.

Table 3. Prevalence of symptomatic dry eye in each region of Thailand.

Region	Symptomatic dry eye (n, %)	
	Without, n=1334 (14.2)	With, n=8031 (85.8)
Northern	82 (6.1)	645 (8.0)
Northeast	96 (7.2)	838 (10.4)
Central (except Bangkok)	272 (20.4)	1703 (21.2)
Bangkok	683 (51.2)	3633 (45.3)
Eastern	85 (6.4)	571 (7.1)
Western	33 (2.5)	163 (2)
Southern	83 (6.2)	478 (6)

Table 4. Prevalence of symptomatic dry eye subgroups in each region of Thailand.

Region	Symptomatic dry eye, n (%)			
	Normal, n=1334 (14.2)	Mild, n=2183 (23.3)	Moderate, n=2024 (21.6)	Severe, n=3824, (40.8)
Northern	82 (11.3)	162 (22.3)	159 (21.9)	324 (44.5)
Northeast	96 (10.3)	186 (19.9)	198 (21.2)	454 (48.6)
Central (except Bangkok)	272 (13.8)	484 (24.5)	426 (21.6)	793 (40.1)
Bangkok	683 (15.8)	1,030 (23.9)	938 (21.7)	1,665 (38.6)
Eastern	85 (12.9)	143 (21.8)	144 (22)	284 (43.3)
Western	33 (16.8)	47 (24)	37 (18.9)	79 (40.3)
Southern	83 (14.8)	131 (23.3)	122 (21.8)	225 (40.1)

Blink Rate Per Minute and Maximum Blink Interval

Blink rate differed significantly between participants with (median 18, IQR 16 blinks) and without (median 16, IQR 16) blinks) symptomatic DE (P<.001). A significant difference was also found in MBI between participants with and without symptomatic DE (median 8.90, IQR 10.80 seconds vs median 8.90, IQR 14.8 seconds, respectively; P<.001). Binary logistic regression showed a significant association between the DE group and blink rate (univariate OR 1.02; 95% CI 1.02-1.03; P<.001) and between the DE group and MBI (univariate OR 0.98; 95% CI 0.98-0.99; P<.001). After controlling for risk factors, including age, sex, VDT use, and educational level, this association was sustained in both blink rate (multivariate-adjusted OR 1.01; 95% CI 1.01-1.02; P<.001) and MBI (multivariate-adjusted OR 0.98; 95% CI 0.98-0.99; P<.001).

Best Spectacle-Corrected Visual Acuity

The BSCVA (logMAR) in users without symptomatic DE (median 0.20, IQR 0.40; mean 0.22, SD 0.21) was significantly different from that of the symptomatic DE group (median 0.20, IQR 0.40; mean0.23, SD 0.21; P=.02). Mean BSCVA was 0.22 (95% CI 0.21-0.23) in users without symptomatic DE; among symptomatic users, the mean was 0.22 (95% CI 0.21-0.23) in mild cases, 0.22 (95% CI 0.21-0.23) in moderate cases, and 0.24 (95% CI 0.24-0.25) in severe cases. Moreover, the median BSCVA in users without symptomatic DE and with mild, moderate, and severe symptomatic DE was 0.2 (IQR 0.4), and there was a statistically significant difference (P<.001). Pairwise comparison revealed a difference between the severe group and each of the three other subgroups, normal (P<.001), mild (P<.001), and moderate (P<.001), but not between the latter three groups (normal vs mild, P=.17; normal vs moderate, P=.26; mild vs moderate; P=.36).

Risk Factors

As shown in Table 5, binary logistic regression found that symptomatic DE was more prevalent in female users (multivariate-adjusted OR 1.83; 95% CI 1.59-2.09; P<.001), those reporting VDT use of >6-8 hours per day (multivariate-adjusted OR 1.59; 95% CI 1.15-2.19; P=.005), and those with an educational level lower than bachelor's degree (multivariate-adjusted OR 1.30; 95% CI 1.03-1.64; P=.02). OSDI scores were significantly higher in female (mean 31.55, SD 17.87) than male (mean 26.53, SD 17.67, P<.001) users, in those with6 hours (mean 32.50, SD 18.27) than those with 6 hours of VDT use (mean 27.59, SD 16.98, P<.001), and in those with an educational level lower than bachelor's degree and bachelor's degree (mean 31.23, SD 17.94, P<.001) versus an educational level higher than bachelor's degree (mean 26.81, SD 17.47). Moreover, the Southern, Western, and Central regions and Bangkok had significantly less impact on symptomatic DE compared with the Northeastern region in binary logistic regression.

Table 5. Risk factors for symptomatic dry eye compared with those without dry eye.

	Univariate OR[a] (95% CI)	*P* value	Multivariate-adjusted OR (95% CI)	*P* value
Age (years)				
15-20	1 [reference]	N/A[b]	1 [reference]	N/A
21-30	0.99 (0.85, 1.18)	.97	1.07 (0.88, 1.29)	.48
31-40	0.68 (0.55, 0.82)	<.001	0.85 (0.67, 1.06)	.15
41-50	0.69 (0.55, 0.88)	.002	0.99 (0.76, 1.28)	.94
>50	0.47 (0.38, 0.58)	<.001	0.77 (0.60, 0.99)	.04
Sex (female vs male)	2.06 (1.81, 2.34)	<.001	1.83 (1.59, 2.09)	<.001
VDT[c] use (hours/day)				
Less than 1 hour	1 [reference]	N/A	1 [reference]	N/A
1-4 hours	1.11 (0.80, 1.53)	.53	1.08 (0.77, 1.50)	.67
>4-6 hours	1.49 (1.10, 2.02)	.01	1.31 (0.96, 1.80)	.09
>6-8 hours	1.87 (1.39, 2.55)	<.001	1.59 (1.15, 2.19)	.005
>8 hours	2.29 (1.69, 3.12)	<.001	1.86 (1.35, 2.58)	<.001
Educational level				
Higher than bachelor's degree	1 [reference]	N/A	1 [reference]	N/A
Lower than bachelor's degree	1.58 (1.31, 1.92)	<.001	1.30 (1.03, 1.64)	.02
Bachelor's degree	1..51 (1.30, 1.76)	<.001	1.18 (0.99, 1.39)	.06
Region				
Northeast	1 [reference]	N/A	1 [reference]	N/A
Northern	0.90 (0.67, 1.23)	.51	0.93 (0.68, 1.27)	.64
Central (except Bangkok)	0.72 (0.56, 0.92)	.008	0.76 (0.59, 0.98)	.03
Bangkok	0.61 (0.49, 0.76)	<.001	0.64 (0.51, 0.81)	<.001
Eastern	0.77 (0.56, 1.05)	.10	0.79 (0.57, 1.08)	.13
Western	0.57 (0.37, 0.87)	.009	0.62 (0.40, 0.97)	.03
Southern	0.66 (0.48, 0.90)	.01	0.68 (0.49, 0.93)	.02

[a]OR: odds ratio

[b]N/A: not applicable.

[c]VDT: visual display terminal.

Discussion

Principal Results

Technology has evolved rapidly in recent years, and smartphones provide one example of this, having transformed dramatically to provide sophisticated communication and data access, including health information [16]. In this study, we used smartphone technology by creating the app "Dry eye or not?" to evaluate the countrywide and regional prevalence of symptomatic DE in Thailand (85.8%). This app enables recording of blink rate (median 18, IQR 16 blinks) and maximum blink interval (median 8.90, IQR 10.80 seconds) associated with DE [14]. Moreover, we used this app to collect BSCVA and demographic data including age, sex, hours of VDT use per day, regions where individuals lived, and educational levels to assess users' relationship with diagnosed symptomatic DE.

The prevalence of symptomatic DE in Thailand in our study was 85.8%, higher than the 34% prevalence reported in 2006 [17] and 14.2% in 2012 [18]. It was also much higher than the 5% to 50% prevalence reported by the Tear Film and Ocular Surface Society's Dry Eye Workshop Study II in 2017 [19], which included prevalence of either or both DE symptoms and signs. However, the symptomatic DE prevalence reported by Inomata et al [6] in 2019 based on data collected using a smartphone app was 74% in Japan. These prevalence values indicate an increasing trend over time [20]. Moreover, Asian race is a known risk factor for DE; consistent with this, prevalence in this study and others including Asian populations is higher than in studies including other races [6,20]. In addition, the fact that both this study and Inomata et al [20] used

smartphone apps for data collection may have biased results toward users with high daily VDT use, another risk factor for DE. In this study, the prevalence of symptomatic DE was highest in the Northeastern region (89.7%), followed by the Northern region (88.7%) of Thailand. The prevalence of severe symptomatic DE was higher than other severity grades in every region of Thailand, and the highest prevalence of severe DE was also in the Northeastern region (48.6%), followed by the Northern region (44.5%). Moreover, binary logistic regression analysis showed that the Northeastern region had a significantly greater impact on symptomatic DE compared with Southern, Western, and Central regions and Bangkok. According to climatic data from the Thai Meteorological Department, the Northern, Northeastern, and Central regions have lower relative humidity than other regions, but the annual average temperature is similar among the regions. High temperature, low humidity, and wind are known risk factors for DE [19,21]. Additionally, air pollutants including ozone (O_3), particulate matter 2.5 (PM 2.5), and sulfur dioxide (SO_2) have been identified as risk factors for DE [22]. Relatively high concentrations of O_3 are found in Central, Northeastern, and Northern regions, high SO_2 concentrations in Central and Northeastern regions [23], and high PM 2.5 concentrations in Bangkok, Central, and Northern regions [24]. High pollution and low humidity may explain the high prevalence of symptomatic DE in Northern and Northeastern regions of Thailand in this study. The Central region also has high pollution and low relative humidity but had a lower prevalence of symptomatic DE in this study. One possible explanation for this is that the population in urban locations in the Central region may be equipped with better health care education, knowledge of the health care system, and access to medication.

Blinking is well established as an associated factor in ocular surface sensation and is commonly quantified by measuring the blink rate or its reciprocal value, such as MBI [14]. This relationship has been demonstrated by many studies that show a link between an increased blink rate in DE and ocular surface irritation, surface dryness, or an unstable tear film, suggesting a blink rate test as a screening tool for DE [25-28]. The rate of spontaneous eye blinking has a complex relationship with ocular surface health, including DE status [28]. In this study, the median spontaneous blink rate (18 blinks per minute) in users with symptomatic DE was significantly higher than in the normal group. The mean blink rate of DE groups in previous studies ranged from 28.55 blinks per minute to 15.32 blinks per 20 seconds and varied considerably, including in this study [25-28]. Interblink interval (IBI) and MBI have both been shown by many studies to be related to DE, with a mean IBI in DE ranging from 2.56 seconds to 12.52 seconds, while the criterion MBI was reported as 12.4 seconds with a sensitivity of 82.5% and a specificity of 51% [14,28,29]. Similarly, the mean MBI in our study was 11.80 seconds with a median of 8.9 seconds. Moreover, according to binary logistic regression, increased blink rate and decreased MBI were associated with symptomatic DE. These findings suggest that blink rate and MBI measured using smartphone technology may be used as screening tools for symptomatic DE, promoting self-diagnosis of symptomatic DE. Because research suggests disagreement between signs and symptoms of DE [30], the efficacy and accuracy of the app developed in this study may be further improved by incorporating factors relating to both signs and symptoms of DE.

DE is a disease associated with ocular surface inflammation, which causes irregularity of the ocular surface and reduced uncorrected visual acuity, the latter having been demonstrated by several studies in patients with DE. Moreover, BSCVA could be reduced in a severe grade of DE. In 2019, Zczotka-Flynn et al [31] also reported that BSCVA was reduced in individuals with worse mean OSDI score [32]. In this study, the BSCVA in the normal group was statistically significantly better than in the symptomatic DE group, and a severe grade of symptomatic DE group had statistically significantly worse BSCVA compared with other groups from the subgroup analysis. The symptomatic DE groups were believed to have instability of tear film layer, which resulted in the irregularity of ocular surface and consequently had a negative impact on optical quality as the air-tear film interface contributes the most to the ocular refractive power [33].

In this study, most participants (6495/9482, 66%) were under 30 years old. The small proportion of older participants in this study is similar to previous research using a smartphone app [6]. Previous studies have shown that the prevalence of DE increases with age [19,20]. However, in this study, participants aged over 50 were less likely to have symptomatic DE than those under 30 years of age. This finding is in accordance with some previous studies [6,34] indicating that older participants may be less likely to report ocular symptoms than younger individuals due to reduced corneal sensitivity in older participants resulting from reductions in corneal density and substance P (a neuropeptide secreted by sensory nerves that modulates nociceptive pain) in older age [35]. Moreover, the proportion of older participants in this study was low, and this relatively small sample of older individuals may have affected prevalence estimates. In addition, younger people tend to report more daily hours of VDT use, and this may have contributed to DE in younger participants.

The female sex has been identified and widely reported as an important factor in DE [19] and was also found to be a risk factor in this study. Sex hormones have roles in tear synthesis. Androgen binding to receptors on the meibomian glands leads to increased lipid synthesis and secretion, while estrogen (predominantly in females) binding lessens lipid production [36].

VDT use has been identified as a risk factor for DE in many studies, but the number of hours of VDT use per day in individuals with DE varies between studies from more than 4 to more than 8 hours a day [6,19,37-39]. In this study, VDT use for more than 6 hours a day was significantly associated with symptomatic DE. The difference in reported periods of VDT use constituting a risk factor for DE may be due to the different criteria for DE diagnosis and different ethnicities included. At present, smartphone and other electronic device displays are increasingly used worldwide, and people, particularly those who are young, spend more time on these devices than was the

case in the past. These factors could lead to an increasing prevalence of DE in the future.

According to the logistic regression analysis, an educational level lower than a bachelor's degree was found to be a significant factor for symptomatic DE in this study. Thus, the percentage of symptomatic DE participants was highest in the lowest educational level group. After further investigation, we found that a lower educational level corresponds to a younger age group. The majority of people with a bachelor's degree or lower had VDT>8 hours per day. The potential reason for the association between lower education and symptomatic DE was that most participants with a lower level of education were aged 15-30 years old, and they had the highest proportion of having 8 hours or more of VDT use. Therefore, this could signify a generational difference, in that a younger age group was associated with a greater VDT use.

Limitations

We acknowledge some limitations in this study, which were similar to those found in other web- or app-based studies. First, our study defined DE diagnosis solely based on the OSDI questionnaire, representing only symptoms of DE. Second, our participants included only those with smart phone capability, thereby restricting our group to younger participants with a relatively high socioeconomic status and education level [12]. Older participants might have a limited ability to use smartphone because of their physical limitations. Third, symptomatic DE individuals were more likely to participate in our project since their interest in alleviating DE may have acted as a motivating factor, whereas those with no DE symptoms may ignore this app [3]. However, this potential problem was alleviated by the large sample size in this study, which recruited participants from diverse geographic regions throughout Thailand. No clinical examination was possible in this remote data collection format; as a result, we tested blink rate and MBI using the smartphone to assess their association with symptomatic DE. Another limitation of this study has to do with the reliability of the blink rate and MBI measurements, since they were carried out in different temperature and humidity conditions; despite this limitation, this study demonstrated the feasibility of blink rate measurement using a smartphone app and showed the link between blink rate and symptomatic DE. Future research incorporating such tests conducted using this app and conventional clinical examination will help improve and validate this convenient screening tool for DE diagnosis. The last limitation was that at the time of this study, there was no published study that validated the OSDI questionnaires in the Thai language; however, that study is currently underway.

Conclusions

According to the results of this crowdsourced study, in which the prevalence of symptomatic DE in Thailand was 85.8%, blink rate, MBI, BSCVA, and risk factors for DE may be evaluated using a smartphone app. Moreover, blink rate and MBI recorded in this way may identify people at risk of symptomatic DE. The Northeastern region of Thailand showed the highest prevalence of symptomatic DE, followed by the Northern region. Increased blink rate, reduced MBI, and reduced BSCVA were associated with symptomatic DE. Younger age was more strongly associated with symptomatic DE than older age. Female sex, more than 6 hours daily VDT use, and a lower education level were also significant risk factors of symptomatic DE. These findings will lead to further research on the use of smartphone app screening tools with high sensitivity and specificity for diagnosis of DE, enabling early diagnosis and treatment of this condition. This approach to screening for DE may aid the development of strategic plans for health care systems in Thailand.

Acknowledgments

We thank Lita Uthaithammarat, MD (Department of Ophthalmology, Faculty of Medicine, Chulalongkorn University) and Yuda Chongpison, PhD (Biostatistics Excellence Center, Research Affairs, Faculty of Medicine, Chulalongkorn University) for conducting data analyses.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Ocular Surface Disease Index questionnaire.
[DOCX File , 14 KB - mhealth_v10i6e31011_app1.docx]

Multimedia Appendix 2
Sensitivity analysis.
[DOCX File , 15 KB - mhealth_v10i6e31011_app2.docx]

References

1. Bron AJ, de Paiva CS, Chauhan SK, Bonini S, Gabison EE, Jain S, et al. TFOS DEWS II pathophysiology report. Ocul Surf 2017 Jul;15(3):438-510. [doi: 10.1016/j.jtos.2017.05.011] [Medline: 28736340]
2. The epidemiology of dry eye disease: report of the Epidemiology Subcommittee of the International Dry Eye WorkShop (2007). Ocul Surf 2007 Apr;5(2):93-107. [doi: 10.1016/s1542-0124(12)70082-4] [Medline: 17508117]

3. Yamanishi R, Uchino M, Kawashima M, Uchino Y, Yokoi N, Tsubota K. Characteristics of Individuals with Dry Eye Symptoms Without Clinical Diagnosis: Analysis of a Web-Based Survey. J Clin Med 2019 May 21;8(5) [FREE Full text] [doi: 10.3390/jcm8050721] [Medline: 31117304]
4. Martínez-Pérez B, de LTI, López-Coronado M. Mobile health applications for the most prevalent conditions by the World Health Organization: review and analysis. J Med Internet Res 2013;15(6):e120 [FREE Full text] [doi: 10.2196/jmir.2600] [Medline: 23770578]
5. Skrzypecki J, Stańska K, Grabska-Liberek I. Patient-oriented mobile applications in ophthalmology. Clin Exp Optom 2019 Mar;102(2):180-183 [FREE Full text] [doi: 10.1111/cxo.12830] [Medline: 30168194]
6. Inomata T, Iwagami M, Nakamura M, Shiang T, Yoshimura Y, Fujimoto K, et al. Characteristics and risk factors associated with diagnosed and undiagnosed symptomatic dry eye using a smartphone application. JAMA Ophthalmol 2020 Jan 01;138(1):58-68 [FREE Full text] [doi: 10.1001/jamaophthalmol.2019.4815] [Medline: 31774457]
7. Uchino M, Kawashima M, Uchino Y, Suzuki N, Mitamura H, Mizuno M, et al. The evaluation of dry eye mobile apps for screening of dry eye disease and educational tear event in Japan. Ocul Surf 2018 Oct;16(4):430-435. [doi: 10.1016/j.jtos.2018.06.002] [Medline: 29883737]
8. Inomata T, Nakamura M, Iwagami M, Midorikawa-Inomata A, Sung J, Fujimoto K, et al. Stratification of individual symptoms of contact lens-associated dry eye using the iphone app dryeyerhythm: crowdsourced cross-sectional study. J Med Internet Res 2020 Jun 26;22(6):e18996 [FREE Full text] [doi: 10.2196/18996] [Medline: 32589162]
9. Eguchi A, Inomata T, Nakamura M, Nagino K, Iwagami M, Sung J, et al. Heterogeneity of eye drop use among symptomatic dry eye individuals in Japan: large-scale crowdsourced research using DryEyeRhythm application. Jpn J Ophthalmol 2021 Mar;65(2):271-281. [doi: 10.1007/s10384-020-00798-1] [Medline: 33411099]
10. Hirosawa K, Inomata T, Sung J, Nakamura M, Okumura Y, Midorikawa-Inomata A, et al. Diagnostic ability of maximum blink interval together with Japanese version of Ocular Surface Disease Index score for dry eye disease. Sci Rep 2020 Oct 22;10(1):18106 [FREE Full text] [doi: 10.1038/s41598-020-75193-4] [Medline: 33093551]
11. Inomata T, Iwagami M, Nakamura M, Shiang T, Fujimoto K, Okumura Y, et al. Association between dry eye and depressive symptoms: Large-scale crowdsourced research using the DryEyeRhythm iPhone application. Ocul Surf 2020 Apr;18(2):312-319. [doi: 10.1016/j.jtos.2020.02.007] [Medline: 32113987]
12. Inomata T, Nakamura M, Iwagami M, Shiang T, Yoshimura Y, Fujimoto K, et al. Risk factors for severe dry eye disease: crowdsourced research using DryEyeRhythm. Ophthalmology 2019 May;126(5):766-768. [doi: 10.1016/j.ophtha.2018.12.013] [Medline: 30550734]
13. Inomata T, Sung J, Nakamura M, Iwagami M, Okumura Y, Iwata N, et al. Using medical big data to develop personalized medicine for dry eye disease. Cornea 2020 Nov;39 Suppl 1:S39-S46. [doi: 10.1097/ICO.0000000000002500] [Medline: 33055549]
14. Inomata T, Iwagami M, Hiratsuka Y, Fujimoto K, Okumura Y, Shiang T, et al. Maximum blink interval is associated with tear film breakup time: A new simple, screening test for dry eye disease. Sci Rep 2018 Sep 07;8(1):13443 [FREE Full text] [doi: 10.1038/s41598-018-31814-7] [Medline: 30194447]
15. Miller KL, Walt JG, Mink DR, Satram-Hoang S, Wilson SE, Perry HD, et al. Minimal clinically important difference for the ocular surface disease index. Arch Ophthalmol 2010 Jan;128(1):94-101. [doi: 10.1001/archophthalmol.2009.356] [Medline: 20065224]
16. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? an evidence review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
17. Lekhanont K, Rojanaporn D, Chuck RS, Vongthongsri A. Prevalence of dry eye in Bangkok, Thailand. Cornea 2006 Dec;25(10):1162-1167. [doi: 10.1097/01.ico.0000244875.92879.1a] [Medline: 17172891]
18. Kasetsuwan N, Gorvanich S, Erjongmanee S, Thienprasiddhi P, Jitapunkul S. Prevalence of dry eyes in elderly Thai population (the Romklao eye study). Asian Biomed 2012;6(6):875-882. [doi: 10.5372/1905-7415.0606.134]
19. Stapleton F, Alves M, Bunya VY, Jalbert I, Lekhanont K, Malet F, et al. TFOS DEWS II Epidemiology Report. Ocul Surf 2017 Jul;15(3):334-365. [doi: 10.1016/j.jtos.2017.05.003] [Medline: 28736337]
20. Dana R, Bradley JL, Guerin A, Pivneva I, Stillman, Evans AM, et al. Estimated prevalence and incidence of dry eye disease based on coding analysis of a large, all-age United States health care system. Am J Ophthalmol 2019 Jun;202:47-54 [FREE Full text] [doi: 10.1016/j.ajo.2019.01.026] [Medline: 30721689]
21. González-García MJ, González-Sáiz A, de la Fuente B, Morilla-Grasa A, Mayo-Iscar A, San-José J, et al. Exposure to a controlled adverse environment impairs the ocular surface of subjects with minimally symptomatic dry eye. Invest Ophthalmol Vis Sci 2007 Sep;48(9):4026-4032. [doi: 10.1167/iovs.06-0817] [Medline: 17724183]
22. Yu D, Deng Q, Wang J, Chang X, Wang S, Yang R, et al. Air pollutants are associated with dry eye disease in urban ophthalmic outpatients: a prevalence study in China. J Transl Med 2019 Feb 15;17(1):46 [FREE Full text] [doi: 10.1186/s12967-019-1794-6] [Medline: 30767763]
23. Guo Y, Li S, Tawatsupa B, Punnasiri K, Jaakkola JJK, Williams G. The association between air pollution and mortality in Thailand. Sci Rep 2014 Jul 01;4:5509 [FREE Full text] [doi: 10.1038/srep05509] [Medline: 24981315]

24. Pinichka C, Makka N, Sukkumnoed D, Chariyalertsak S, Inchai P, Bundhamcharoen K. Burden of disease attributed to ambient air pollution in Thailand: A GIS-based approach. PLoS One 2017;12(12):e0189909 [FREE Full text] [doi: 10.1371/journal.pone.0189909] [Medline: 29267319]
25. Nakamori K, Odawara M, Nakajima T, Mizutani T, Tsubota K. Blinking is controlled primarily by ocular surface conditions. Am J Ophthalmol 1997 Jul;124(1):24-30. [doi: 10.1016/s0002-9394(14)71639-3] [Medline: 9222228]
26. Tsubota K, Hata S, Okusawa Y, Egami F, Ohtsuki T, Nakamori K. Quantitative videographic analysis of blinking in normal subjects and patients with dry eye. Arch Ophthalmol 1996 Jun;114(6):715-720. [doi: 10.1001/archopht.1996.01100130707012] [Medline: 8639084]
27. Ousler GW, Abelson MB, Johnston PR, Rodriguez J, Lane K, Smith LM. Blink patterns and lid-contact times in dry-eye and normal subjects. Clin Ophthalmol 2014;8:869-874 [FREE Full text] [doi: 10.2147/OPTH.S56783] [Medline: 24833893]
28. Su Y, Liang Q, Su G, Wang N, Baudouin C, Labbé A. Spontaneous eye blink patterns in dry eye: clinical correlations. Invest Ophthalmol Vis Sci 2018 Oct 01;59(12):5149-5156. [doi: 10.1167/iovs.18-24690] [Medline: 30372746]
29. Johnston PR, Rodriguez J, Lane KJ, Ousler G, Abelson MB. The interblink interval in normal and dry eye subjects. Clin Ophthalmol 2013;7:253-259 [FREE Full text] [doi: 10.2147/OPTH.S39104] [Medline: 23403736]
30. Bartlett JD, Keith MS, Sudharshan L, Snedecor SJ. Associations between signs and symptoms of dry eye disease: a systematic review. Clin Ophthalmol 2015;9:1719-1730 [FREE Full text] [doi: 10.2147/OPTH.S89700] [Medline: 26396495]
31. Szczotka-Flynn LB, Maguire MG, Ying G, Lin MC, Bunya VY, Dana R, Dry Eye AssessmentManagement (DREAM) Study Research Group. Impact of dry eye on visual acuity and contrast sensitivity: dry eye assessment and management study. Optom Vis Sci 2019 Jun;96(6):387-396 [FREE Full text] [doi: 10.1097/OPX.0000000000001387] [Medline: 31116166]
32. Rieger G. The importance of the precorneal tear film for the quality of optical imaging. Br J Ophthalmol 1992 Mar;76(3):157-158 [FREE Full text] [doi: 10.1136/bjo.76.3.157] [Medline: 1540559]
33. Benítez-Del-Castillo J, Labetoulle M, Baudouin C, Rolando M, Akova YA, Aragona P, et al. Visual acuity and quality of life in dry eye disease: Proceedings of the OCEAN group meeting. Ocul Surf 2017 Apr;15(2):169-178. [doi: 10.1016/j.jtos.2016.11.003] [Medline: 27913232]
34. Uchino M, Schaumberg DA, Dogru M, Uchino Y, Fukagawa K, Shimmura S, et al. Prevalence of dry eye disease among Japanese visual display terminal users. Ophthalmology 2008 Nov;115(11):1982-1988. [doi: 10.1016/j.ophtha.2008.06.022] [Medline: 18708259]
35. Marco B, Alessandro R, Philippe F, Fabio B, Paolo R, Giulio F. The effect of aging on nerve morphology and substance p expression in mouse and human corneas. Invest Ophthalmol Vis Sci 2018 Nov 01;59(13):5329-5335. [doi: 10.1167/iovs.18-24707] [Medline: 30398624]
36. Peck T, Olsakovsky L, Aggarwal S. Dry eye syndrome in menopause and perimenopausal age group. J Midlife Health 2017;8(2):51-54 [FREE Full text] [doi: 10.4103/jmh.JMH_41_17] [Medline: 28706404]
37. Rossi GCM, Scudeller L, Bettio F, Pasinetti GM, Bianchi PE. Prevalence of dry eye in video display terminal users: a cross-sectional Caucasian study in Italy. Int Ophthalmol 2019 Jun;39(6):1315-1322. [doi: 10.1007/s10792-018-0947-6] [Medline: 29881936]
38. Uchino M, Yokoi N, Uchino Y, Dogru M, Kawashima M, Komuro A, et al. Prevalence of dry eye disease and its risk factors in visual display terminal users: the Osaka study. Am J Ophthalmol 2013 Oct;156(4):759-766. [doi: 10.1016/j.ajo.2013.05.040] [Medline: 23891330]
39. Christensen MA, Bettencourt L, Kaye L, Moturu ST, Nguyen KT, Olgin JE, et al. Direct measurements of smartphone screen-time: relationships with demographics and sleep. PLoS One 2016;11(11):e0165331 [FREE Full text] [doi: 10.1371/journal.pone.0165331] [Medline: 27829040]

Abbreviations

API: application programming interface
BSCVA: best spectacle-corrected visual acuity
DE: dry eye
IBI: interblink interval
IQR: interquartile range
MBI: maximum blink interval
ML: machine learning
O_3: ozone
OSDI: Ocular Surface Disease Index
PM 2.5: particulate matter 2.5
SO_2: sulfur dioxide
VDT: visual display terminal

Edited by L Buis, A Mavragani; submitted 06.06.21; peer-reviewed by T Inomata, L Guo; comments to author 13.07.21; revised version received 10.08.21; accepted 09.05.22; published 22.06.22.

Please cite as:
Kasetsuwan N, Suwan-Apichon O, Lekhanont K, Chuckpaiwong V, Reinprayoon U, Chantra S, Puangsricharern V, Pariyakanok L, Prabhasawat P, Tesavibul N, Chaidaroon W, Tananuvat N, Hirunpat C, Prakairungthong N, Sansanayudh W, Chirapapaisan C, Phrueksaudomchai P
Assessing the Risk Factors For Diagnosed Symptomatic Dry Eye Using a Smartphone App: Cross-sectional Study
JMIR Mhealth Uhealth 2022;10(6):e31011
URL: https://mhealth.jmir.org/2022/6/e31011
doi:10.2196/31011
PMID:35731569

Original Paper

Monitoring Symptoms of COVID-19: Review of Mobile Apps

Suzanna Schmeelk[1,2], EdD, MBA, MS, DPS; Alison Davis[1,3], MS; Qiaozheng Li[1,4], MS; Caroline Shippey[1], MS; Michelle Utah[1], MS; Annie Myers[1], MA; Meghan Reading Turchioe[1], RN, MPH, PhD; Ruth Masterson Creber[1], MSc, RN, PhD

[1]Weill Cornell Medicine, New York, NY, United States

[2]St John's University, Queens, NY, United States

[3]Montefiore Medical Center, Bronx, NY, United States

[4]Memorial Sloan Kettering Cancer Center, New York, NY, United States

Corresponding Author:
Ruth Masterson Creber, MSc, RN, PhD
Weill Cornell Medicine
425 E 61st Street
New York, NY, 10065
United States
Phone: 1 9144340731
Email: rmc2009@med.cornell.edu

Abstract

Background: Mobile health (mHealth) apps have facilitated symptom monitoring of COVID-19 symptoms globally and have been used to share data with health care professionals and support disease prediction, prevention, management, diagnostics, and improvements in treatments and patient education.

Objective: The aim of this review is to evaluate the quality and functionality of COVID-19 mHealth apps that support tracking acute and long-term symptoms of COVID-19.

Methods: We systematically reviewed commercially available mHealth apps for COVID-19 symptom monitoring by searching Google Play and Apple iTunes using search terms such as “COVID-19,” “Coronavirus,” and “COVID-19 and symptoms.” All apps underwent three rounds of screening. The final apps were independently assessed using the Mobile Application Rating Scale (MARS), an informatics functionality scoring system, and the Center for Disease Control and World Health Organization symptom guidelines. The MARS is a 19-item standardized tool to evaluate the quality of mHealth apps on engagement, functionality, aesthetics, and information quality. Functionality was quantified across the following criteria: inform, instruct, record (collect, share, evaluate, and intervene), display, guide, remind or alert, and communicate. Interrater reliability between the reviewers was calculated.

Results: A total of 1017 mobile apps were reviewed, and 20 (2%) met the inclusion criteria. The majority of the 20 included apps (n=18, 90%) were designed to track acute COVID-19 symptoms, and only 2 (10%) addressed long-term symptoms. Overall, the apps scored high on quality, with an overall MARS rating of 3.89 out of 5, and the highest domain score for functionality (4.2). The most common functionality among all apps was the instruct function (n=19, 95%). The most common symptoms included in the apps for tracking were fever and dry cough (n=18, 90%), aches and pains (n=17, 85%), difficulty breathing (n=17, 85%), tiredness, sore throat, headache, loss of taste or smell (n=16, 80%), and diarrhea (n=15, 75%). Only 2 (10%) apps specifically tracked long-term symptoms of COVID-19. The top 4 rated apps overall were state-specific apps developed and deployed for public use.

Conclusions: Overall, mHealth apps designed to monitor symptoms of COVID-19 were of high quality, but the majority of apps focused almost exclusively on acute symptoms. Future apps should also incorporate monitoring long-term symptoms of COVID-19 and evidence-based educational materials; they should also include a feature that would allow patients to communicate their symptoms to specific caregivers or their own health care team. App developers should also follow updated technical and clinical guidelines from the Center for Disease Control and the World Health Organization.

(JMIR Mhealth Uhealth 2022;10(6):e36065) doi:10.2196/36065

KEYWORDS

COVID-19; mobile apps; mobile health; mHealth; symptom assessment; symptom tracking; public health; mobile health application; surveillance; digital surveillance; monitoring system; digital health

Introduction

Monitoring and tracking acute short-term symptoms are important to help identify the combination of symptoms that occur in individuals with COVID-19 for personal and public health purposes [1]. Information gathered from public symptom tracking can also help guide recommendations for self-isolation and testing and help prevent further spread of the virus [2]. The Center for Disease Control (CDC) recommends daily monitoring for symptoms of COVID-19 illness to reduce transmission risk [1], and to holistically understand its full impact.

In addition to short-term symptoms, there is a growing awareness of the long-term symptoms of COVID-19 including fatigue or loss of taste and smell, as well as multiorgan effects on the heart, lungs, renal function, and cognitive functions [3]. The population prevalence of long COVID is unknown but is estimated to affect between 1 in 5 people (symptoms beyond 5 weeks) and 1 in 10 people (symptoms 12+ weeks) [4] who have been infected with COVID-19. Improved understanding of long-COVID symptoms can help health care professionals recognize the most common long-term impacts and add urgency to the public health messaging focused on the importance of COVID-19 vaccination efforts [5].

Since the start of the pandemic, mobile health (mHealth) [6] apps have been leveraged in several ways to control the spread of COVID-19 [7]. These mHealth apps can be used to monitor symptoms, share data with providers, and support disease prediction, prevention, management, diagnostics, and improvements in treatments and patient education, ultimately giving patients more control [8]. The majority of adults in the United States have access to a smartphone (>85%) [9]; thus, mHealth apps are poised to provide scalable and cost-effective delivery of health care at the point of need for patients with COVID-19. The incorporation of pertinent epidemiological and geographic data on the presence of transmittable diseases in a region allows the tracing of cases, which can be used as a successful tool to control the spread of the infection [10]. The mHealth apps can also help solve several COVID-19–related challenges by increasing the reach of reliable information to both patients and health care professionals. Additionally, mobile apps can assist in tracking physical and mental health symptoms, support home monitoring, and reduce the burden of hospitals [7].

The aim of this review was to evaluate the quality and functionality of mHealth apps that support tracking acute and long-term symptoms of COVID-19. Our research focuses on the CDC [1,3,11] and World Health Organization (WHO) [12] symptom guidelines tracked by the apps as medical understandings continue to develop over time. Our analysis can support health care professionals by identifying appropriate mHealth apps for patients regarding COVID-19 acute and long-term symptoms. This review also identified key areas in the quality, functionality, and content of existing COVID-19 apps that can be improved in current and future related app development.

Methods

Systematic Search Criteria and Selection

In June 2021, we systematically searched 2 major app stores: the Apple App Store and Android Google Play Store. In the 2 app stores, search terms included “COVID-19,” “Coronavirus,” “COVID-19 and Symptoms,” “Coronavirus and Symptoms,” “COVID-19 and Symptom Monitoring,” and “Coronavirus and Symptom Monitoring” to identify relevant apps. Following subsequent searches of “COVID-19” and “Coronavirus,” the authors found that the results were the same. The terms “COVID-19,” “Coronavirus,” “COVID-19 and Symptoms,” and “COVID-19 and Symptom Monitoring” were then collapsed to streamline the search.

All apps underwent three rounds of screening. During the first round, the title and screenshots of the apps were used to exclude those that were exclusively for contact tracing (without symptom tracking functionality), were not available in English, required an institution-specific login, were clinical guidelines for clinicians, were games, or provided general or other COVID-19–related information. During the second round of review, screenshots and descriptions of the apps were reviewed to exclude apps that were too general, only had a singular function (eg, only track and trace), or were not relevant to tracking COVID-19 symptoms. During the third round of review, the final apps were downloaded, and apps requiring institutional credentialing were excluded.

Evaluation Measures or Rating Tools

All apps were evaluated using the Mobile Application Rating Scale (MARS) [13,14], IQVIA for Healthcare Informatics [15] functionality scoring system, and specific COVID-19 symptoms according to the CDC [1,3,11] and WHO COVID-19 health topic guidelines [12]. Both the MARS and IQVIA functionality scores were used for this review. The MARS functionality score focuses on performance, ease of use, navigation, and gestural design of the app [13]. The 7 IQVIA functionality scores focus on the scope of functions, including informing, instructing, recording, displaying, guiding, reminding, and communicating information [14].

The MARS [13] is a widely used, multidimensional tool to evaluate the quality of mobile health apps and was developed based on semantic analysis and a combination of relevant literature. The MARS was used to rate app quality and includes four sections: classification, quality, satisfaction, and a modifiable app-specific section. The classification section provides descriptive information about the apps. The objective app quality section includes 19 items divided into four scales: engagement, functionality, aesthetics, and information quality. MARS items are scored using a 5-point Likert scale (1=inadequate, 2=poor, 3=acceptable, 4=good, and 5=excellent)

[13]. The final MARS scores include a total mean score for the engagement, functionality, aesthetics, and information subscales. In addition, the MARS also includes a subjective quality score and an app-specific subscale that assesses perceived effect on the user's knowledge, attitudes, and intentions to change, as well as the likelihood of changing the identified targeted behaviors. The functionality domain measures whether an app is easy to learn, encourages seamless navigation, supports a logical flow, and examines the overall app gestural design. The engagement score includes evaluation of whether the app is fun, interesting, customizable, and interactive (eg, sends alerts, messages, reminders, and feedback, and enables sharing), and is well targeted to the intended audience.

The IQVIA functionality score (formerly termed the IMS functionally score) is based on 7 functionality criteria and 4 functional subcategories as described in detail in the IQVIA Institute for Healthcare Informatics report [14]. IQVIA is not an acronym; the name represents a merger between two companies, IMS Health and Quintiles. The apps were evaluated on the 7 functionality criteria of inform, instruct, record, display, guide, remind or alert, and communicate. The record functionality criteria were further scored on 4 subcategories: collect (ability to enter or store data on individual phone), share (ability to transmit data), evaluate (ability to evaluate health data by caregiver or health care entity), and intervene (ability to alert in response to collected data or propose behavioral intervention).

The specific symptoms tracked in each app were evaluated against acute and long-term symptoms of COVID-19 as reported by the CDC [1] and the WHO [12] as of June 2021. Acute COVID-19 symptoms included fever, dry cough, tiredness, aches and pains, sore throat, diarrhea, conjunctivitis, headache, loss of taste or smell, skin rash or discoloration of fingers or toes, difficulty breathing or shortness of breath, chest pain or pressure, and loss of speech or movement. Long-COVID-19 symptoms, defined as more than 4 weeks from the initial infection [3], included tiredness or fatigue, difficulty thinking or concentrating ("brain fog"), headache, loss of smell or taste, dizziness while standing, heart palpitations, chest pain, difficulty breathing or shortness of breath, cough, joint or muscle pain, depression or anxiety, fever, and symptoms that worsen after physical or mental activities. Symptoms that are consistent with both short-term and long-term COVID-19 (eg, cough), were evaluated based on how data entry was described in the app and the current version of the app.

Data Extraction and Data Analysis

We created a Google data extraction form consisting of questions from (1) the MARS questionnaire [13], (2) IQVIA functionality guidelines [14], and (3) the CDC [1] and WHO [12] COVID-19 symptoms. A total of 5 reviewers independently evaluated the 4 randomly selected apps using the data extraction form to assess interrater reliability, which was acceptable (0.75-0.83). Domains with low agreement between reviewers (<0.7) were discussed until consensus was reached, apps were rereviewed, and interrater reliability was recalculated. The remaining apps were independently evaluated by 2 reviewers. Mean MARS scores for each domain, MARS total scores, and IQVIA functionality scores are reported.

Results

The Android Google Play and Apple App Store searches identified 1017 potentially relevant apps, of which 20 (2%) met our final inclusion criteria. The search strategy (Figure 1) shows the number of apps included and excluded in each round of review. In total, 50% (n=509) of the apps had a government affiliation, 10% (n=102), were affiliated with a university and the rest (n=406, 40%) were developed by private companies (Table 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the screening process. MARS: Mobile Application Rating Scale.

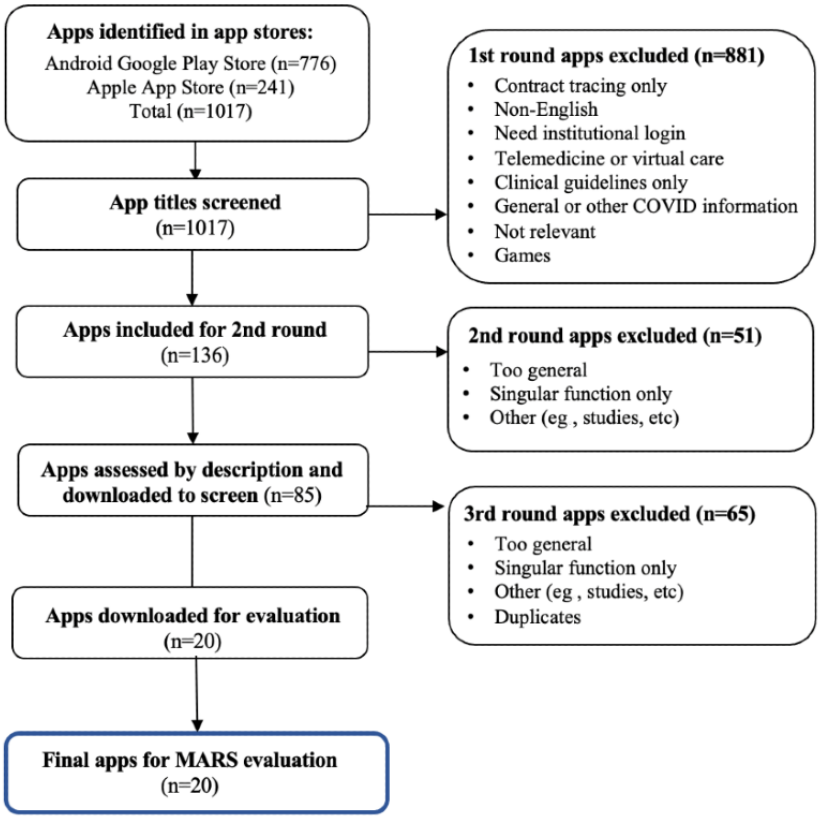


Table 1. Descriptive characteristics of the apps included.

App name	Star rating	N ratings	Current version	Last update	Affiliation	Geographic location–dependent?	Privacy policy
Apollo	4.8	16	1.2.8	2020	Commercial[a]	No	Yes
Beebe Covid-19 Screening Tool	3	2	N/A[b]	N/A	Commercial[a]	No	Yes
Geohealthapp Covid19 Tracker	3.6	8	1.2	2020	Commercial[a]	No	Yes
Healthy Together - Covid-19	4.9	71,219	1.5.8	2021	Commercial[a]	No	Yes
HowWefeel	4.8	54,879	1.13.3	2020	Commercial[a]	No	Yes
Patientsphere Cv	2.8	9	1.1	2020	Commercial[a]	No	Yes
Apple Covid-19	4.2	2782	5	2020	Commercial[a]	No	Yes
Bc Covid-19 Support	3.4	7	1.41	2021	Government	Yes	Yes
Care 19 Diary	3.8	591	3.6	2020	Government	Yes	Yes
Covid Alert DE	4.2	30	1.2.2	2020	Government	Yes	Yes
Covid Alert NJ	3.7	301	1.1.4	2021	Government	Yes	Yes
Covid Alert NY	4.6	544	1.1.7	2021	Government	Yes	Yes
Covid Alert Pennsylvania	4.1	301	2.0.0	2021	Government	Yes	Yes
Covid Coach	4.8	722	1.6	2021	Government	No	Yes
Covid Trace Nevada	2.8	172	1.2.16	2021	Government	Yes	Yes
Crush Covid RI	3.6	195	3	2020	Government	Yes	Yes
Soco Covid-19 Check	2.3	50	1.1.2	2020	Government	Yes	Yes
Check Covid	4.5	74	3	2021	University	No	Yes
My Covid-19 Tracker	3.8	5	1.2.2	2021	University	No	No
Canada Covid-19	4.4	16	5.17	2021	Commercial[a]	Yes	Yes

[a]For-profit.

[b]N/A: not applicable.

Mobile Application Rating Scale

The average overall MARS rating for the 20 apps was 3.89 with the functionality domain having the highest score (4.20) and the engagement score having the lowest average score (3.52). The top 4 rated apps overall were state-specific apps developed and deployed for public use in New York, Pennsylvania, Rhode Island, and New Jersey (Table 2), between July and October 2020.

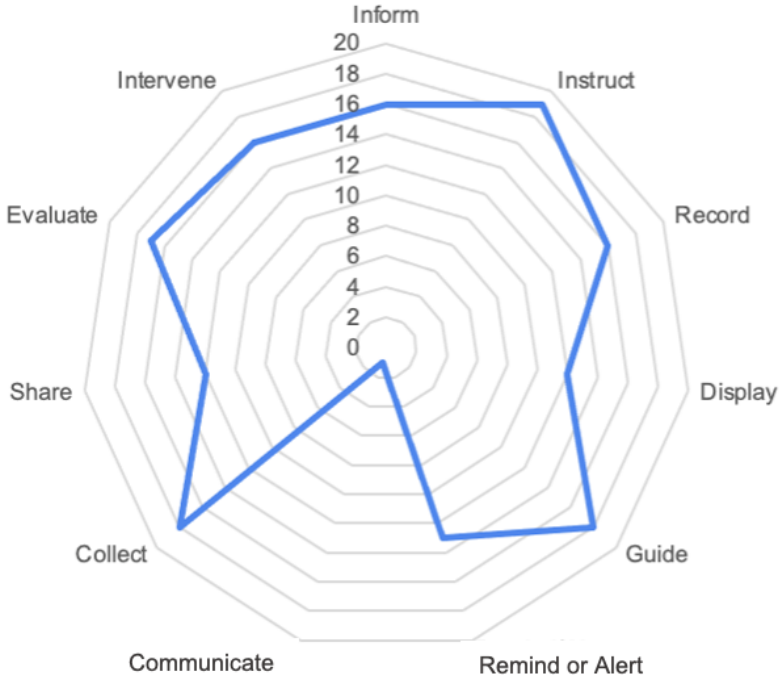
Table 2. Mobile Application Rating Scale (MARS) quality scores.

App name	Engagement	Functionality	Aesthetics	Information	Overall score
Covid Alert NY	4.30	4.50	4.83	4.00	4.41
Covid Alert Pennsylvania	4.00	4.13	4.83	4.00	4.24
Crush Covid RI	4.05	4.44	4.42	4.04	4.23
Covid Alert NJ	4.00	4.50	4.50	3.86	4.21
HowWeFeel	3.90	4.55	4.47	3.80	4.18
Covid Coach	3.90	4.63	4.67	3.43	4.16
Healthy Together: Covid-19	3.92	4.40	4.47	3.74	4.13
Check Covid	4.00	4.63	4.33	3.29	4.06
Canada Covid-19	3.60	4.50	4.33	3.71	4.04
Covid Alert DE	3.60	4.63	4.17	3.64	4.01
Apple Covid-19	3.60	4.33	4.22	3.74	3.97
SoCo Covid-19 Check	3.60	4.25	4.00	3.79	3.91
My Covid-19 Tracker	3.30	3.88	4.00	3.64	3.70
Apollo Covid-19	2.95	4.00	4.17	3.64	3.69
Beebe Covid-19 Screening	3.10	4.00	3.67	3.64	3.60
Care19 Diary	2.90	3.50	4.00	3.71	3.53
Patientsphere CV	3.10	4.00	3.50	3.43	3.51
BC Covid-19 Support	3.12	3.85	3.53	3.43	3.48
Covid Trace Nevada	3.00	3.58	3.67	3.57	3.46
Geohealthapp Covid19 Tracker	2.40	3.75	3.83	2.79	3.19
Average score	3.52	4.20	4.18	3.64	3.89

IQVIA Functionality

Overall, the “instruct” function was the most common among all apps (n=19, 95%) (Figure 2). In many of the apps, the instruct function aided in directing users on next steps based on the symptoms that were recorded. The “Beebe COVID-19” screening app was the only app that had the ability to communicate between patients and health care providers. There were no apps that had all 11 functionalities. Another domain that the majority of the apps had was the collect function (n=18, 90%). These apps collected patients’ daily COVID-19 symptoms and collected some demographic information (n=18, 90%) including: age, name, date of birth, ethnicity, and geographical location.

Figure 2. Functionality scores based on IQVIA for Healthcare Informatics.



COVID-19 WHO and CDC Guidelines

The acute COVID-19 symptoms tracked in each app (Table 3), and long-term symptoms (Table 4) were evaluated against the recommendations provided by the CDC [1] and the WHO [12]. The majority of apps focused on acute COVID-19 symptoms (n=18) with the most common acute symptoms including fever and dry cough (n=18, 90%), aches and pains (n=17, 85%), difficulty breathing (n=17, 85%), tiredness, sore throat, headache, loss of taste or smell (n=16, 80%), and diarrhea (n=15, 75%).

Table 3. COVID-19 acute symptoms.

App name	Fever	Dry cough	Tiredness	Aches and pains	Sore throat	Diarrhea	Conjunctivitis	Headache	Loss of taste or smell	Skin rash	Difficulty breathing or shortness of breath	Chest pain or pressure
Apollo	✓[a]	✓	✓	✓	✓	✓	—[b]	✓	✓	—	✓	—
Apple Covid-19	✓	✓	✓	✓	✓	✓	—	✓	✓	—	✓	✓
BC COVID-19 Support	✓	✓	✓	✓	✓	—	—	✓	✓	✓	✓	✓
Beebe COVID-19 Screening Tool	✓	✓	—	✓	✓	✓	—	✓	✓	✓	✓	—
Canada COVID-19	✓	✓	✓	✓	✓	✓	—	✓	✓	✓	✓	—
Care19 Diary	✓	✓	✓	—	—	—	—	—	—	—	✓	—
Check Covid	✓	✓	✓	✓	✓	✓	—	✓	✓	—	✓	✓
COVID Alert DE	✓	✓	✓	✓	✓	✓	—	✓	✓	—	✓	—
COVID Alert NJ	✓	✓	✓	✓	✓	✓	—	✓	✓	—	✓	✓
COVID Alert NY	✓	✓	✓	✓	✓	✓	—	✓	✓	—	✓	—
COVID Alert Pennsylvania	✓	✓	✓	✓	✓	✓	—	✓	✓	—	✓	—
COVID Trace Nevada	✓	✓	✓	✓	✓	✓	—	✓	✓	—	✓	—
Crush Covid RI	✓	✓	✓	✓	✓	✓	—	✓	✓	—	✓	—
Healthy Together	✓	✓	✓	✓	✓	✓	—	✓	✓	—	✓	—
HowWeFeel	✓	✓	✓	✓	✓	✓	—	✓	✓	—	✓	✓
My COVID-19 Tracker	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
PatientSphere	✓	✓	✓	✓	✓	—	—	—	—	—	—	—
SoCo COVID-19 Check	✓	✓	—	✓	—	✓	—	✓	✓	—	✓	—

[a]✓ implies that the symptom is measured in the app.

[b]— implies that the symptoms are not measured in the app.

Table 4. COVID-19 long-term symptoms.

App name	Dry cough	Tiredness	Headache	Loss of taste or smell	Symptoms that worsen after physical or mental activities	Difficulty breathing or shortness of breath	Chest pain or pressure	Depression or anxiety
BC COVID-19 Support	✓[a]	✓	✓	✓	✓	✓	✓	—[b]
HowWeFeel	—	—	—	—	—	—	—	✓

[a]✓ implies that the symptoms are measured in the app.

[b]— implies that the symptoms are not measured in the app.

Discussion

Principal Findings

The aim of this review was to evaluate COVID-19 mHealth apps that support tracking acute and long-term symptoms of COVID-19. The majority of the apps reviewed focused on acute symptoms, including fever and dry cough, aches and pains, difficulty breathing, tiredness, sore throat, headache, loss of taste or smell, and diarrhea. Few apps measured long-term symptoms of COVID-19. Overall, the apps scored high on quality, and the most common functionality was providing the user with instructions on what to do in response to symptoms.

Previous reviews of mHealth apps for COVID-19 were published by Alanzi [16] and Davalbhakta [17]. The review published by Alanzi [16] reviewed apps from 7 countries (Saudi Arabia, India, Singapore, Australia, Italy, United Kingdom, and the United States). These apps were reviewed on functionality (eg, purpose, services offered, and networking technologies) and effectiveness (eg, learnability, communication strategies, and design). There were no overlapping apps between the review published by Alanzi [16] and this review. The review by Davalbhakta and colleagues [17] was consistent methodologically by using the MARS to review the apps, in addition to adding more functionalities specific to COVID-19 (eg, individual tracking, contact tracing, and health care worker training). The review by Davalbhakta [17] and this review shared 6 common apps (Apollo, PatientSphere, Apple COVID-19, Bc COVID-19 Support, Check COVID, and Canada COVID-19). Other reviews did not evaluate short- and long-term

symptom monitoring according to the CDC [1] or WHO COVID-19 health topic guidelines [12].

The highest quality apps were developed by individual states within the United States to support acute COVID-19 symptom tracking. Overall, the state-specific apps had the most up-to-date and accurate information, which could be attributed to collaborative efforts between local, county, or state departments of health [18-23]. The state-specific apps all listed most of the common symptoms and provided information about local COVID-19 case counts and developments in research. Strengths of the state-specific apps included long-term storage of the symptom evaluations, high-quality aesthetics, comprehensive symptom lists, and evidence-based information that was aligned with CDC guidelines [1]. While the contact tracing functions were likely most useful to in-state residents, the symptom tracking functionalities were available for anyone regardless of their home state.

At the time of app development and deployment, less was known about long-term symptoms of COVID-19 [24]. Only 2 (10%) apps specifically tracked long-term symptoms of COVID-19 (Table 4), though the symptoms of long COVID-19 are increasingly concerning [25-27]. As such, mHealth apps that track long COVID-19 symptoms could play a significant role in helping to manage them more efficiently and gather additional data about how this disease is affecting patients over a long period of time [25-27]. The authors suggest that app developers should follow updated technical and clinical guidelines from the CDC and the WHO to ensure consistency and efficacy of long-term symptom monitoring. Future research should focus on expanding these apps to support patients with long-term symptoms [3] and providing educational materials about the disease.

The largest gap in app functionality was in the communication feature, which could be further developed to allow patients to communicate their symptoms to chosen family members for supportive care or their own health care professionals. These algorithms are useful for providing basic information about what to do in response to the symptoms and quarantine guidance. However, many patients, especially those with complex care needs, require additional support and communication with their health team to manage long-term symptoms of COVID-19. There are specific legal and regulatory issues around data sharing that are likely the reasons why many apps do not actively support more data-sharing features. Additional areas of future development include expansion into pediatric populations and long-term symptom tracking integrated into mHealth apps. Lastly, these apps should be evaluated for usability and inclusivity characteristics, including non-English languages.

Limitations

A limitation of this review is that it provides a snapshot of the app landscape at a specific point in time when the app stores have continued to rapidly evolve. Many of the apps are also going on and off the market based on the introduction of novel variants. To mitigate the limitation of subjectivity of reviewing apps, we applied a rigorous multistep methodology including using the MARS [13,28,29] and IQVIA for Healthcare Informatics [15] functionality scoring system and mapping all the apps to the CDC and WHO COVID-19 health topic guidelines for symptom monitoring.

Conclusion

In general, mHealth apps for tracking COVID-19 symptoms offer a potential solution to help people identify and track virus symptoms. Overall, the mHealth apps had high quality, considering the expedited needs during the COVID-19 pandemic. Future apps should also incorporate monitoring long-term symptoms of COVID-19 and include a communication feature that could allow patients to communicate their symptoms to specific caregivers or their own health care team.

Acknowledgments

This work was funded by R00NR016275 and R00NR016275-05S1 (PI: Masterson Creber).

Conflicts of Interest

None declared.

References

1. Coronavirus Disease 2019 (COVID-19). Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/index.html [accessed 2021-11-10]
2. Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med 2020 Jul 11;26(7):1037-1040 [FREE Full text] [doi: 10.1038/s41591-020-0916-2] [Medline: 32393804]
3. Long COVID or Post-COVID Conditions. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects.html [accessed 2021-11-10]
4. Michelen M, Manoharan L, Elkheir N, Cheng V, Dagens A, Hastie C, et al. Characterising long COVID: a living systematic review. BMJ Glob Health 2021 Sep 27;6(9):e005427 [FREE Full text] [doi: 10.1136/bmjgh-2021-005427] [Medline: 34580069]

5. Wanga V, Chevinsky JR, Dimitrov LV, Gerdes ME, Whitfield GP, Bonacci RA, et al. Long-Term Symptoms Among Adults Tested for SARS-CoV-2 - United States, January 2020-April 2021. MMWR Morb Mortal Wkly Rep 2021 Sep 10;70(36):1235-1241 [FREE Full text] [doi: 10.15585/mmwr.mm7036a1] [Medline: 34499626]
6. Park Y. Emerging New Era of Mobile Health Technologies. Healthc Inform Res 2016 Oct;22(4):253-254 [FREE Full text] [doi: 10.4258/hir.2016.22.4.253] [Medline: 27895955]
7. Kondylakis H, Katehakis DG, Kouroubali A, Logothetidis F, Triantafyllidis A, Kalamaras I, et al. COVID-19 Mobile Apps: A Systematic Review of the Literature. J Med Internet Res 2020 Dec 09;22(12):e23170 [FREE Full text] [doi: 10.2196/23170] [Medline: 33197234]
8. Istepanian R, AlAnzi T. Chapter Twenty Two - Mobile health (m-health): Evidence-based progress or scientific retrogression. In: Biomedical Information Technology (Second Edition). London, England: Academic Press; 2020:717-733.
9. Mobile Vendor Market Share United States Of America: Apr 2021 - Apr 2022. Stat Counter. URL: https://gs.statcounter.com/vendor-market-share/mobile/united-states-of-america [accessed 2021-11-10]
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. Science Brief: Options to Reduce Quarantine for Contacts of Persons with SARS-CoV-2 Infection Using Symptom Monitoring and Diagnostic Testing. Centers for Disease Control and Prevention.: CDC COVID-19 Science Briefs. Centers for Disease Control and Prevention (US) URL: https://tinyurl.com/mtjy9j7u [accessed 2022-05-26]
12. Coronavirus disease (COVID-19). World Health Organization. URL: https://www.who.int/health-topics/coronavirus [accessed 2021-11-10]
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. Aitken M. Patient apps for improved healthcare from novelty to mainstream. IMS Institute. 2013. URL: http://ignacioriesgo.es/wp-content/uploads/2014/03/iihi_patient_apps_report_editora_39_2_1.pdf [accessed 2022-05-26]
15. Powering healthcare with connected intelligence. IQVIA. URL: https://www.iqvia.com/ [accessed 2022-02-17]
16. Alanzi T. A Review of Mobile Applications Available in the App and Google Play Stores Used During the COVID-19 Outbreak. JMDH 2021 Jan;Volume 14:45-57. [doi: 10.2147/jmdh.s285014]
17. 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]
18. Crush COVID. Rhode Island Department of Health. URL: https://covid.ri.gov/covid-19-prevention/crush-covid-ri-app [accessed 2021-11-10]
19. COVID Alert NY. New York State. URL: https://coronavirus.health.ny.gov/covid-alert-ny [accessed 2021-11-10]
20. COVID Alert NJ. New Jersey COVID-19 Information Hub. URL: https://COVID-19.nj.gov/pages/app [accessed 2021-11-10]
21. COVID Alert DE. Delaware.gov. URL: https://coronavirus.delaware.gov/covidalert/ [accessed 2021-11-10]
22. COVID Trace. Nevada Health Response. URL: https://nvhealthresponse.nv.gov/covidtrace/ [accessed 2021-11-10]
23. COVID Alert PA App. Commonwealth of Pennsylvania. 2020 Aug 18. URL: https://www.pa.gov/covid/covid-alert-pa/ [accessed 2021-11-10]
24. Del Rio C, Collins LF, Malani P. Long-term Health Consequences of COVID-19. JAMA 2020 Nov 03;324(17):1723-1724 [FREE Full text] [doi: 10.1001/jama.2020.19719] [Medline: 33031513]
25. Wang F, Kream RM, Stefano GB. Long-Term Respiratory and Neurological Sequelae of COVID-19. Med Sci Monit 2020 Nov 01;26:e928996-1-e928996-10. [doi: 10.12659/msm.928996]
26. Yelin D, Wirtheim E, Vetter P, Kalil AC, Bruchfeld J, Runold M, et al. Long-term consequences of COVID-19: research needs. The Lancet Infectious Diseases 2020 Oct;20(10):1115-1117. [doi: 10.1016/s1473-3099(20)30701-5]
27. Kamal M, Abo Omirah M, Hussein A, Saeed H. Assessment and characterisation of post-COVID-19 manifestations. Int J Clin Pract 2021 Mar 03;75(3):e13746 [FREE Full text] [doi: 10.1111/ijcp.13746] [Medline: 32991035]
28. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016 Mar 10;13(1):35 [FREE Full text] [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
29. Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and Evaluation of Mindfulness-Based iPhone Apps. JMIR Mhealth Uhealth 2015 Aug 19;3(3):e82 [FREE Full text] [doi: 10.2196/mhealth.4328] [Medline: 26290327]

Abbreviations

CDC: Center for Disease Control
MARS: Mobile Application Rating Scale
mHealth: mobile health

WHO: World Health Organization

Edited by L Buis; submitted 12.01.22; peer-reviewed by D Katehakis, SS Amritphale, X Wu, S Pit, A Joseph; comments to author 08.02.22; revised version received 01.03.22; accepted 03.05.22; published 01.06.22.

Please cite as:
Schmeelk S, Davis A, Li Q, Shippey C, Utah M, Myers A, Reading Turchioe M, Masterson Creber R
Monitoring Symptoms of COVID-19: Review of Mobile Apps
JMIR Mhealth Uhealth 2022;10(6):e36065
URL: https://mhealth.jmir.org/2022/6/e36065
doi:10.2196/36065
PMID:35609313

Original Paper

Assessment of Smartphone Apps for Common Neurologic Conditions (Headache, Insomnia, and Pain): Cross-sectional Study

Mia T Minen[1], MPH, MD; Alexis George[1], BA; Erica Camacho[2], MS; Leslie Yao[3]; Ananya Sahu[3]; Maya Campbell[3]; Mia Soviero[3]; Quazi Hossain[4]; Deepti Verma[4]; John Torous[2], MBI, MD

[1]Department of Neurology, New York University Langone Health, New York, NY, United States

[2]Department of Psychiatry, Beth Israel Deaconess Medical Center, New York, NY, United States

[3]Barnard College, New York, NY, United States

[4]The City College of New York, New York, NY, United States

Corresponding Author:
Mia T Minen, MPH, MD
Department of Neurology
New York University Langone Health
222 East 41st Street
New York, NY, 10017
United States
Phone: 1 2122637744
Email: minenmd@gmail.com

Abstract

Background: There are thousands of apps for individuals struggling with headache, insomnia, and pain, but it is difficult to establish which of these apps are best suited for patients' specific needs. If clinicians were to have access to a platform that would allow them to make an informed decision on the efficacy and feasibility of smartphone apps for patient care, they would feel confident in prescribing specific apps.

Objective: We sought to evaluate the quality of apps for some of the top common, disabling neurologic conditions (headache, insomnia, and pain) based on principles derived from the American Psychiatric Association's (APA) app evaluation model.

Methods: We used the Mobile Health Index and Navigation database and expanded upon the database's current supported conditions by adding 177 new app entries. Each app was rated for consistency with the APA's app evaluation model, which includes 105 objective questions based on the following 5 major classes of consideration: (1) accessibility, (2) privacy and security, (3) clinical foundation, (4) engagement style, and (5) interoperability. These characteristics were evaluated to gain a broader understanding of the significant features of each app category in comparison against a control group.

Results: Approximately 90% (187/201) of all apps evaluated were free to download, but only 50% (63/201) of headache- and pain-related apps were truly free. Most (87/106, 81%) sleep apps were not truly free to use. The apps had similar limitations with limited privacy, accessibility, and crisis management resources. For example, only 17% (35/201) of the apps were available in Spanish. The apps offered mostly self-help tools with little tailoring; symptom tracking was the most common feature in headache- (32/48, 67%) and pain-related apps (21/47, 45%), whereas mindfulness was the most common feature in sleep-related apps (73/106, 69%).

Conclusions: Although there are many apps for headache, pain, and insomnia, all 3 types of apps have room for improvement around accessibility and privacy. Pain and headache apps share many common features, whereas insomnia apps offer mostly mindfulness-based resources. Given the many available apps to pick from, clinicians and patients should seek apps that offer the highest-quality features, such as complete privacy, remedial features, and the ability to download the app at no cost. These results suggest that there are many opportunities for the improvement of apps centered on headache, insomnia, and pain.

(JMIR Mhealth Uhealth 2022;10(6):e36761) doi:10.2196/36761

KEYWORDS

headache; pain; insomnia; mobile health; smartphone apps; mobile phone

Introduction

There is a health crisis in the United States whereby people cannot access neurologic care in a timely manner [1]. As smartphones and digital tools increase in popularity, with 85% ownership as of 2021, compared to only 35% in 2011 [2], many mobile health tools have been developed as a means to provide self-management and other strategies to patients. This is especially true for common and disabling neurologic conditions such as headache, sleep, and pain disorders [3]. With 1 in every 6 American adults experiencing migraine and severe headache [4], 70 million American adults experiencing sleep problems [5], and 1 in 5 American adults experiencing chronic pain [6], there is a clear need for treatment. Although a quick search in an app store for "headache," "pain," or "sleep" may reveal countless apps, the apps listed at the top of a search result do not necessarily offer benefits in terms of utility and efficacy compared to others [7]. There have been many reviews of mental health apps in each of the app store marketplaces (ie, Apple iTunes and Google Play) [8], but there have been fewer reviews for neurology-focused apps [9]. Given the inherent risks of apps, including privacy concerns [10-12], mixed evidence around efficacy, and broad usability concerns [13,14], clinicians and patients need to be aware of the state of these public-facing apps and be able to understand their risks and benefits.

Despite the broad risks in the digital health space, emerging evidence suggests the potential benefits of apps for neurological conditions. Even simple headache tracking apps have been shown to help with the management of symptoms [15]. In randomized controlled trials, apps for insomnia have shown benefits such as significant reduction in sleep-related impairment of quality of life and mental well-being [16]. Presently, apps for pain management are expanding in scope, with features such as pain impact recording and medication tracking [17]. To aid users in their efforts to discover apps that are accessible, safe, effective, and evidence based, several app libraries have been developed. One such publicly available tool that considers these and many other metrics when evaluating an app is the Mobile Health Index and Navigation Database (MIND) [8,18]. MIND is the largest open and publicly accessible database of mental health and neurology-focused apps—with over 600 apps, each rated across 105 criteria and updated at least every 6 months. Per recent research, MIND is unique as it also represents diversity, equity, and inclusion criteria, such as accessibility features and language options, which offer a more comprehensive window into apps utility [19]. To understand the current state of mobile apps for neurological disorders, using headache and pain as the leading causes of nonfatal health loss [20] and insomnia as the common sleep disorder [21], we applied the 105 metrics found in MIND to the top neurological apps discoverable on iOS and Android devices.

We sought to (1) search app stores for headache-, pain-, and sleep-related apps and review them using the MIND database, and (2) evaluate the characteristics of these apps and compare them to a control group of apps across unique features not yet reported in the literature, including language (Spanish), crisis management, and the ability to connect with providers. We predict that if clinicians were to have access to a platform that would allow them to make an informed decision on the efficacy and feasibility of smartphone apps for patient care, they would feel confident in prescribing specific apps.

Methods

App Selection

This study used the MIND database, published by the Division of Digital Psychiatry at Beth Israel Deaconess Medical Center [18]. Details about MIND have been published previously [8,22,23]. In brief, MIND is the largest publicly available database of mental health apps; it currently logs 656 mental health apps in the commercial market across a variety of supported conditions and acts as an open resource for users to filter apps by features of their personal preference. To add an app to the MIND database, there are 105 objective questions that are answered based on the following 5 major classes of consideration within the American Psychiatric Association's app evaluation model: accessibility, privacy and security, clinical foundation, engagement style, and interoperability (Multimedia Appendix 1). All data in the MIND database are publicly accessible through the MINDapps website [18]. A screenshot of the MIND database is provided in Figure 1.

The database does not include apps for which the cost of download exceeds US $10 or those not accessible to the public. Prior to this study, the database had 24 apps for pain-, headache-, or sleep-related conditions. The study expanded upon the current database by adding 177 new app entries for the supported conditions—47 apps for pain, 48 for headache, and 106 for sleep—accounting for overlap between categories, amounting to a total of 201 apps analyzed in this study. The selection of new apps was conducted as follows: to gain an understanding of the app marketplace for headache-, pain-, and sleep-related apps, terms such as "headache," "pain," and "insomnia" were searched in both the iOS and Google Play stores in June and July 2021. The first 50 apps that appeared on each platform (Apple App Store and Google Play Store) for each search term (headache, pain, and sleep) were compiled into a Google Sheet. Thus, 300 apps were discovered in our preliminary search. Some of these apps appeared within the first 50 apps searched on both platforms. Apps were assessed for relevance to the neurological conditions of interest. Given that there is no standard definition for these apps and many apps appearing in a search may not be related to wellness or health (eg, a gaming app), all apps selected for evaluation were agreed upon for relevance via consensus of all raters, all of whom are authors. Apps were excluded if they were irrelevant, clinician facing, nonfunctional, unavailable in English, and required an access code to use them. Upon removal of apps that did not meet the inclusion criteria, the remaining 177 relevant apps were evaluated and added to the MIND database. Figure 2 shows a flowchart detailing app selection and app rating process.

Given that all data were entered into the MIND, unique to this study, we sought to publicly share all app evaluations, so others can expand on these results and use this raw data to explore and find relevant apps today, as well as to aid neurologists in making informed decisions around choosing smartphone apps for patients with these conditions.

Figure 1. Screenshot of the main page of the MINDapps database taken in April 2022 (the screenshot was taken after the study was completed).

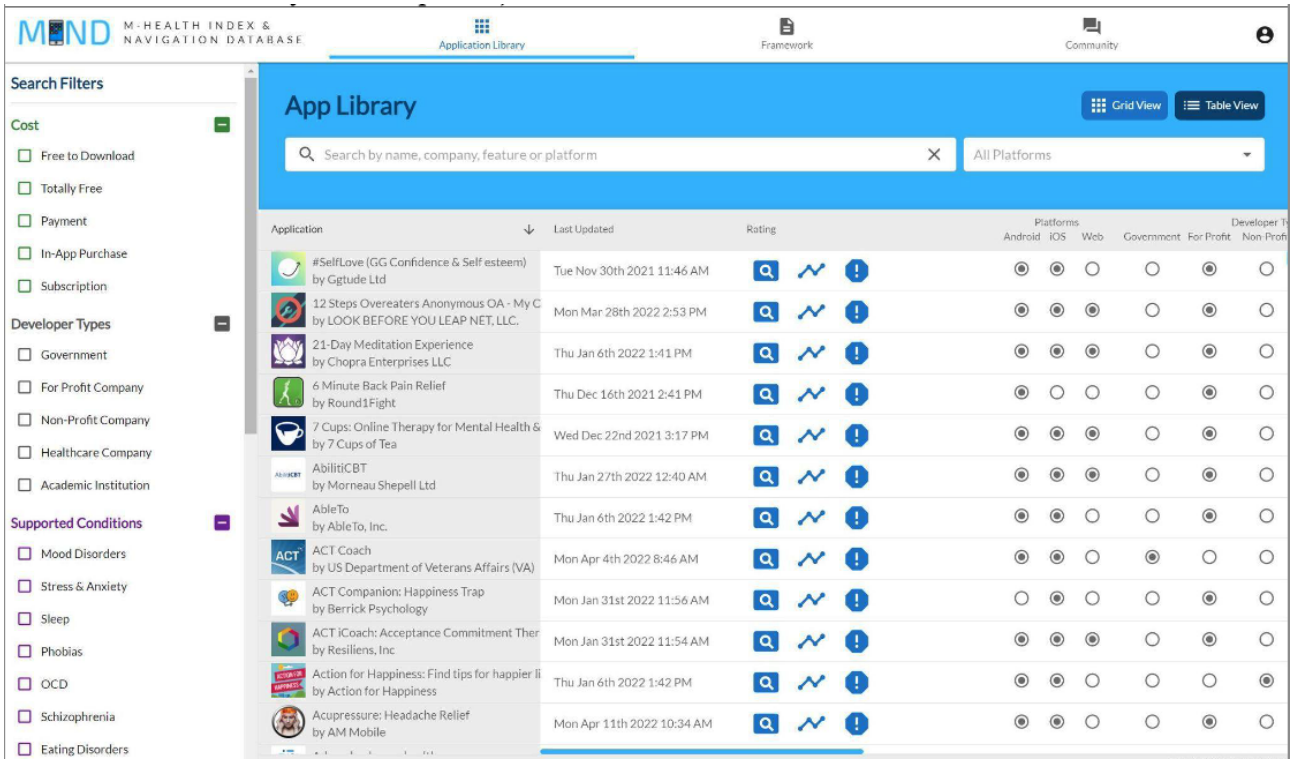
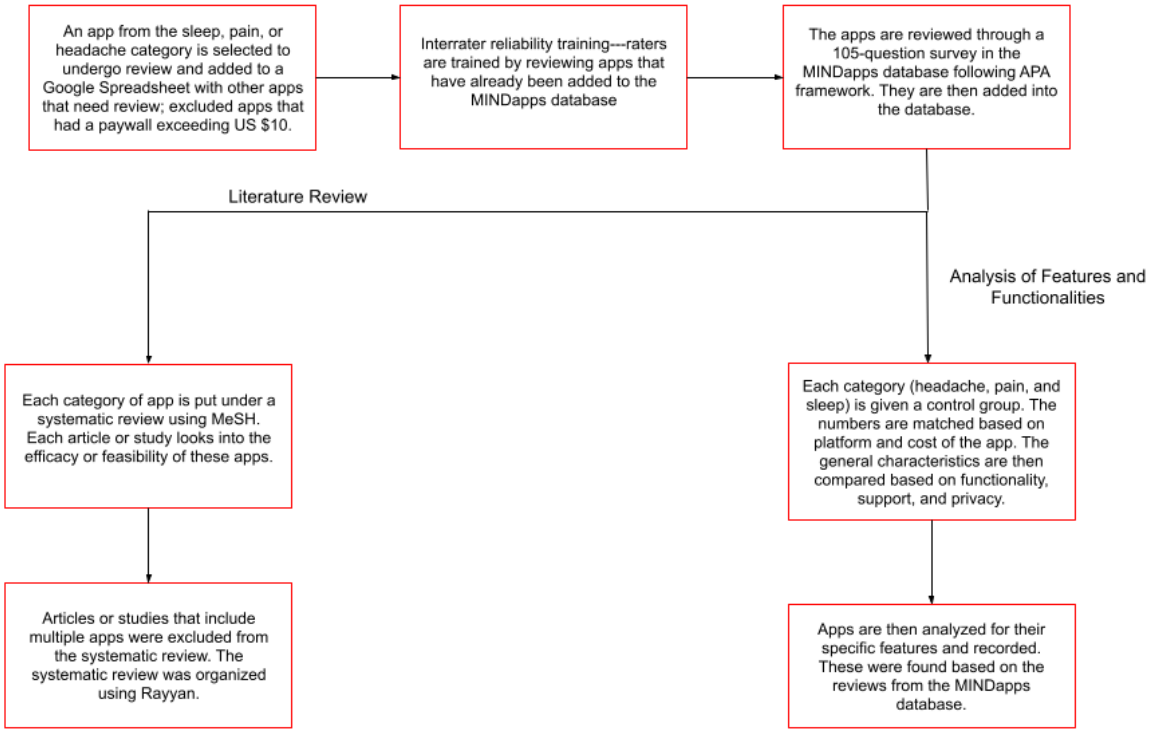


Figure 2. Flowchart detailing app selection and app rating process. APA: American Psychiatric Association; MeSH: Medical Subject Headings.



Statistical Analysis

A total of 6 app raters underwent interrater reliability training [23], and 1 app rater evaluated each app. Interrater reliability was assessed using Cohen κ statistic, for which raters demonstrated very good interrater reliability (defined as a κ value above 0.750), with an average κ value of 0.859 across all apps rated. Discrepancies between the raters were addressed individually through discussion and subsequently resolved by clarifying any discrepancy in the description of each question.

As this was purely an exploratory study, we used Excel (Microsoft Corporation) and reported descriptive statistics.

Results

Overview

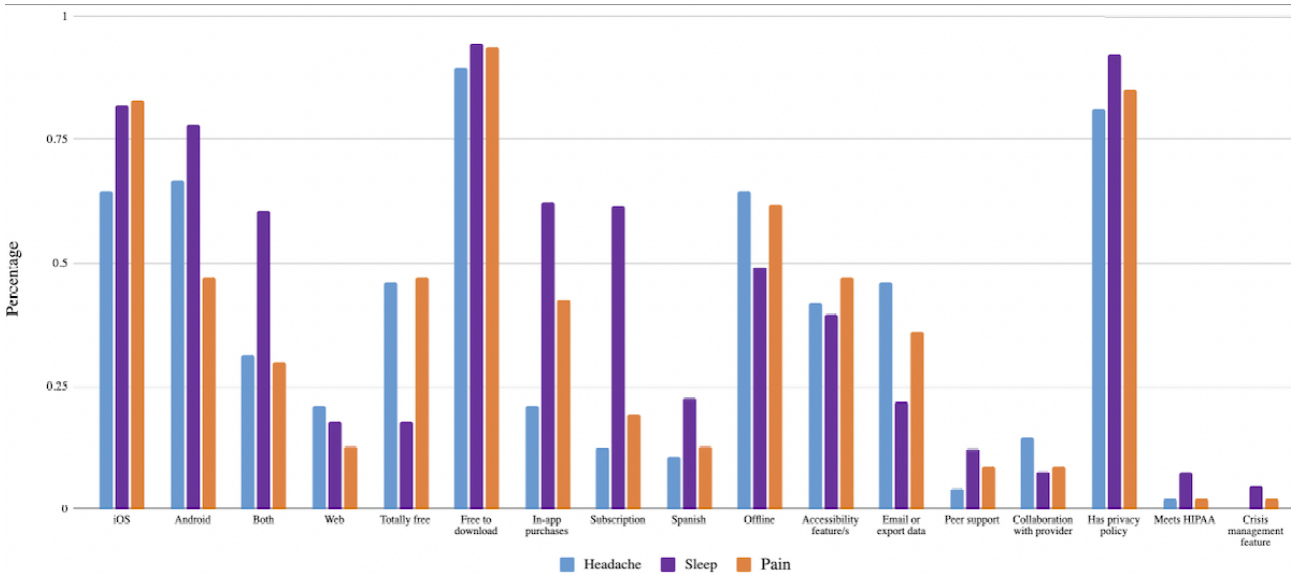
Of the 656 apps available in MIND in July 2021, a total of 201 were related to headache, pain, or insomnia. Overall, these 201 apps focusing on headache, pain, and insomnia offered common features for tracking symptoms, tracking medication, journaling, and psychoeducation. Very few apps (26/201, 13%) used biological data, defined as metrics obtained from external devices (eg, wearables or built-in phone sensors), to monitor personal health. Examples of biodata collected include skin conductance, heart rate, and sleep quality. Across all 3 types of apps examined, we found similarities in terms of platform cost, special features (eg, Spanish language and accessibility features), clinician support, and privacy features as shown in both Table 1 and Figure 3.

Table 1. General characteristics of headache and migraine apps, sleep and insomnia apps, and pain-related apps (N=201).

Characteristics	Apps, n (%)		
	Headache and migraine (n=48)	Sleep and insomnia (n=106)	Chronic pain (n=47)
Platforms			
iOS	31 (65)	87 (82)	39 (83)
Android	32 (65)	83 (78)	22 (47)
Both iOS and Android	15 (31)	64 (60)	14 (30)
Web	10 (21)	19 (18)	6 (13)
Cost			
Totally free	22 (46)	19 (18)	22 (47)
Free to download	43 (90)	100 (94)	44 (94)
In-app purchases	17 (21)	66 (62)	20 (43)
Subscription	10 (13)	65 (61)	9 (19)
Functionality			
Spanish	5 (11)	24 (23)	6 (13)
Offline	31 (65)	52 (49)	29 (62)
Accessibility features	20 (42)	42 (40)	22 (47)
Email or export data	25 (46)	23 (22)	17 (36)
Support			
Peer support	2 (4)	13 (12)	4 (9)
Collaboration with provider	7 (15)	8 (8)	4 (9)
Privacy			
Includes privacy policy	39 (8)	98 (92)	40 (85)
Meets HIPAA[a] requirements	1 (2)	5 (5)	1 (2)
Crisis management feature	0 (0)	8 (8)	1 (2)

[a]HIPAA: Health Insurance Portability and Accountability Act.

Figure 3. Characteristics and features of headache, pain, and sleep apps. HIPPA: Health Insurance Portability and Accountability Act.



Accessibility

Apps were accessible on Apple, Android, and web browsers, although less than 50% (63/201) were truly free of cost. Over 90% (187/201) of apps in all categories were free to download, but this did not guarantee no-cost or even low-cost use. Over 50% (112/201) of apps across all disease states offered functionalities for working offline, that is, without an internet connection. Approximately 17% (35/201) of these apps supported Spanish.

Crisis Management and Privacy

Apps in all 3 categories demonstrated lack of crisis management features, with 0% (0/48) of headache-related apps, 2% (1/47) of pain-related apps, and 8% (8/106) of sleep-related apps offering crisis resources in terms of providing resources for a hotline or contact with a medical professional. Most apps for pain, headache, or insomnia did offer a privacy policy, with 88% (177/201) of apps among all categories containing information on user data storage and usage. Although apps that are not part of health care accountability organizations are not subject to the Health Insurance Portability and Accountability Act (HIPAA), 3% (7/201) of apps stated that they were HIPAA compliant.

Self-help and Hybrid Use With a Clinician

Most apps were self-help centered, but some included collaboration of a clinician (either from the app or outside the app). This was offered by 15% (7/46) of headache apps, 9% (8/47) of pain apps, and 8% (4/106) of sleep apps. Proportions of apps offering peer support were similarly low.

Overall Functionality

The reported functionality offered by these apps is shown in Table 2. Results show that headache and pain apps shared many common features, with tracking symptoms as the most used feature and mindfulness as one of the least used features. In contrast, the apps that focused on sleep had mindfulness as their most common feature and symptom tracking as one of the least common features.

Table 2. Top features for headache and migraine apps, sleep and insomnia apps, and pain-related apps (N=201).

Apps and features	Values, n (%)
Headache and migraine apps (n=48)	
Track symptoms	32 (67)
Track medication	30 (63)
Journaling	16 (33)
Psychoeducation	14 (29)
Track sleep	6 (13)
Track mood	6 (13)
Physical health	5 (11)
Mindfulness	5 (11)
Biodata	5 (11)
Sleep and insomnia apps (n=106)	
Mindfulness	73 (69)
Deep breathing	58 (55)
Psychoeducation	32 (30)
iCBT[a] or sleep therapy	26 (25)
Track mood	26 (25)
Goal settings or habits	25 (24)
Journaling	25 (24)
Track sleep	24 (23)
Physical health	15 (14)
Pain apps (n=47)	
Track symptoms	21 (45)
Physical health	20 (43)
Track medication	16 (34)
Psychoeducation	15 (32)
Physical health exercises	12 (26)
Journaling	10 (21)
Track sleep	8 (17)
Track mood	8 (17)
Mindfulness	6 (13)

[a]iCBT: internet-based cognitive behavioral therapy.

Other Considerations

Each app was evaluated across 105 individual criteria, and all results are publicly accessible and searchable today through the MINDapps database [18]. Figure 1 shows an example of how a reader can interactively explore and search these apps across individual questions. We do not provide scores for certain categories (privacy, functionality, etc), as MINDapps allows users to select their own filters and create their own criteria based on personal needs.

Discussion

Principal Findings

Our review of apps for headache, insomnia, and pain is the largest review of publicly available neurology-focused offerings to date, with results derived from over 200 apps each categorized across 105 dimensions. Pain and headache apps share many common features, whereas insomnia apps offer mostly mindfulness-based resources. We found that apps mostly offered self-help tools with little tailoring, and that symptom tracking was the most common feature in headache- (32/48, 67%) and pain-related apps (21/47, 45%), whereas mindfulness was the most common feature in sleep-related apps (73/106, 69%).

Despite the number of apps and 3 unique conditions, we found numerous commonalities, including limited privacy, accessibility, and crisis management resources, in these mostly self-help tools. In terms of features offered, tracking and mindfulness-related features were most common, with individual apps offering varied ratios or types of these core features. These results suggest opportunities for innovation around the structure of apps themselves, as well as how they deliver tracking or mindfulness, with any innovation presenting transdiagnostic benefits. The numerous overlapping features offered by these apps also suggest that clinicians and patients today can be demanding in selecting an app, as there are likely minimal differences in their core functionality. Using MINDapps [18], they can explore which apps may offer the exact app features desired.

In selecting apps beyond core functions, our results highlight concerns about the structure of apps in terms of privacy, accessibility, and use. Although it is well known that most health-related apps had privacy and access issues, our results are novel for apps in the neurology field. A June 2021 review of 20,911 Android apps across the entire digital health space found that 28.1% of apps offer no privacy policy [24], and our results showed this to be the case in only 9% (177/201). This lower proportion is encouraging but may also be due to only including apps that appeared to be clinically relevant. On September 15, 2021, the Federal Trade Commission (FTC) noted that for wellness apps not covered by HIPAA, the FTC will now expect them to follow HIPAA-related rules around breaches, suggesting that apps will need to offer a change in the required security process [25].

A recent review of mental health–focused apps found that almost 15% supported Spanish, which is consistent with our result of 17% (35/201) for the neurology-focused apps we reviewed [26]. This result suggests an immediate opportunity to increase reach, while better supporting diversity and inclusion. Most apps we reviewed were self-help focused, with only a small fraction designed to be used in partnership with a peer or clinicians. Across the broader digital health field, there is growing evidence that apps used in partnership with others may be more engaging and effective than the self-help ones. Lessons already learned about low engagement with mental health apps [27] may help these neurology-related apps develop as more engaging relationship-based tools that could offer more support for hybrid use.

Comparison With Prior Work

Our findings about the common features underlying apps for headache, insomnia, and pain suggest room for transdiagnostic innovation around all these apps. The common features of symptom tracking, and mindfulness are also the most common features in mental health apps [8], suggesting a potential synergy between these two fields. This result makes sense as apps offer a practical platform to deliver behavioral-based treatments and remotely capture symptoms, which are themselves shared aspects between psychiatry and neurology. Although the insomnia apps focused more on behavioral interventions, the headache and pain apps focused more on symptom tracking.

Furthermore, although smartphone apps may show promising results for patient care, it remains challenging to evaluate their effectiveness. Clinical studies do not often feature a valid control condition or simulate the challenges of real-world app engagement, which is frequently low [27]. Thus, the MIND framework does not rate the quality of scientific evidence for apps nor their engagement, given the lack of consensus or data availability around these points. Our results are therefore best interpreted as signals of what these apps are claiming to offer, with the recommendations for personal use and exploration with the app itself to determine whether the feature meets the needs for each clinical use case. Clinicians should feel empowered to go to MINDapps [18] and search for apps that they feel would meet the standards for their patient's needs. The results of this paper can help calibrate expectations and guide clinicians in searching for apps that meet the unique demands of each patient served.

Limitations

Although each app was downloaded and tested, some aspects of the data coded within this study are based on the description of the app itself, such as who the developer is and what the privacy policy reports. As a result, coded information could be inaccurate. Although app descriptions may hold biases, an advantage of this approach is that apps are constantly being updated, and it is feasible and practical to update apps on the website regularly. Apps were reviewed by only 1 rater. It was also found that there were several apps that falsely advertised their features; other times, the majority of the features were locked behind a paywall, so they could not be seen or used unless the premium version was paid for. Although this study represents perhaps the largest analysis of neurology-related apps, there remains no simple way to identify all relevant apps, given the limitations in searching function on the Apple and Android marketplaces.

Conclusion

Although the number of headache, sleep, and pain apps on the market continues to expand, there are numerous opportunities for content improvement. Many of these apps were lacking in privacy, accessibility, and crisis management resources—features that would significantly improve the app platforms. The results of this study suggest an opportunity for improvement in app structure and the delivery of important features. Patient care may be improved with the incorporation of a transdiagnostic approach to health-based smartphone apps.

Authors' Contributions

MTM designed and conceptualized the study, drafted the manuscript, as well as revised it. AG conducted data analysis and interpretation, drafted the manuscript, and revised it. EC conducted data analysis and interpretation, drafted and revised the

manuscript, and was responsible for the acquisition of data. LY, AS, QH, DV, MS, and MC contributed to the acquisition, analysis, and interpretation of data.

Conflicts of Interest

MTM is supported by NCCIH (grant K23 AT009706-05). MTM also helped develop the RELAXaHEAD app, which has shared IP between New York University Langone Health and Irody Inc. The app was not used in this study, as it has been used as a research app to date and has not been available to the public. JT developed the MIND database with support from the Argosy Foundation.

Multimedia Appendix 1
App review questions.
[DOCX File , 27 KB - mhealth_v10i6e36761_app1.docx]

References

1. Dorsey ER, George BP, Leff B, Willis AW. The coming crisis: obtaining care for the growing burden of neurodegenerative conditions. Neurology 2013 Apr 24;80(21):1989-1996. [doi: 10.1212/wnl.0b013e318293e2ce] [Medline: 23616157]
2. Mobile fact sheet. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2022-06-01]
3. Minen MT, Stieglitz EJ. Wearables for neurologic conditions. Neurol Clin Pract 2020 Sep 28;11(4):e537-e543. [doi: 10.1212/cpj.0000000000000971] [Medline: 34484952]
4. Burch R, Rizzoli P, Loder E. The prevalence and impact of migraine and severe headache in the united states: figures and trends from government health studies. Headache 2018 Apr 12;58(4):496-505. [doi: 10.1111/head.13281] [Medline: 29527677]
5. Hafner M, Stepanek M, Taylor J, Troxel WM, Van Stolk C. Why sleep matters—the economic costs of insufficient sleep: a cross-country comparative analysis. RAND corporation. 2017. URL: https://www.rand.org/content/dam/rand/pubs/research_reports/RR1700/RR1791/RAND_RR1791.pdf [accessed 2022-06-01]
6. Zelaya CE, Dahlhamer JM, Lucas JW, Connor EM. Chronic pain and high-impact chronic pain among U.S. adults, 2019. NCHS Data Brief 2020 Nov(390):1-8. [Medline: 33151145]
7. Wisniewski H, Liu G, Henson P, Vaidyam A, Hajratalli NK, Onnela J, et al. Understanding the quality, effectiveness and attributes of top-rated smartphone health apps. Evid Based Ment Health 2019 Feb 11;22(1):4-9 [FREE Full text] [doi: 10.1136/ebmental-2018-300069] [Medline: 30635262]
8. Lagan S, D'Mello R, Vaidyam A, Bilden R, Torous J. Assessing mental health apps marketplaces with objective metrics from 29,190 data points from 278 apps. Acta Psychiatr Scand 2021 Aug 29;144(2):201-210. [doi: 10.1111/acps.13306] [Medline: 33835483]
9. Minen MT, Gopal A, Sahyoun G, Stieglitz E, Torous J. The functionality, evidence, and privacy issues around smartphone apps for the top neuropsychiatric conditions. J Neuropsychiatry Clin Neurosci 2021;33(1):72-79 [FREE Full text] [doi: 10.1176/appi.neuropsych.19120353] [Medline: 32669020]
10. Benjumea J, Ropero J, Rivera-Romero O, Dorronzoro-Zubiete E, Carrasco A. Privacy assessment in mobile health apps: scoping review. JMIR Mhealth Uhealth 2020 Jul 02;8(7):e18868 [FREE Full text] [doi: 10.2196/18868] [Medline: 32459640]
11. Minen MT, Stieglitz EJ, Sciortino R, Torous J. Privacy issues in smartphone applications: an analysis of headache/migraine applications. Headache 2018 Jul 04;58(7):1014-1027 [FREE Full text] [doi: 10.1111/head.13341] [Medline: 29974470]
12. Minen MT, Gopal A, Sahyoun G, Stieglitz E, Torous J. The functionality, evidence, and privacy issues around smartphone apps for the top neuropsychiatric conditions. J Neuropsychiatry Clin Neurosci 2021 Jan;33(1):72-79 [FREE Full text] [doi: 10.1176/appi.neuropsych.19120353] [Medline: 32669020]
13. Molloy A, Anderson PL. Engagement with mobile health interventions for depression: a systematic review. Internet Interv 2021 Dec;26:100454 [FREE Full text] [doi: 10.1016/j.invent.2021.100454] [Medline: 34621626]
14. Friis-Healy EA, Nagy GA, Kollins SH. It is time to REACT: opportunities for digital mental health apps to reduce mental health disparities in racially and ethnically minoritized groups. JMIR Ment Health 2021 Jan 26;8(1):e25456 [FREE Full text] [doi: 10.2196/25456] [Medline: 33406050]
15. Bensink M, Shah S, Shah N, Desai P, Khan F, Rubin A, et al. Tracking migraine digitally: the future of migraine management. J Nurse Pract 2021 Apr;17(4):462-470. [doi: 10.1016/j.nurpra.2021.01.014]
16. Espie CA, Emsley R, Kyle SD, Gordon C, Drake CL, Siriwardena AN, et al. Effect of digital cognitive behavioral therapy for insomnia on health, psychological well-being, and sleep-related quality of life: a randomized clinical trial. JAMA Psychiatry 2019 Jan 01;76(1):21-30 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.2745] [Medline: 30264137]
17. 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(106):1-10. [doi: 10.1186/s12911-019-0827-7]
18. MIND: M-Health Index and Navigation Database. URL: https://mindapps.org/ [accessed 2022-06-01]

19. Ramos G, Ponting C, Labao JP, Sobowale K. Considerations of diversity, equity, and inclusion in mental health apps: a scoping review of evaluation frameworks. Behav Res Ther 2021 Dec;147:103990 [FREE Full text] [doi: 10.1016/j.brat.2021.103990] [Medline: 34715396]
20. GBD 2017 US Neurological Disorders Collaborators, Feigin VL, Vos T, Alahdab F, Amit AML, Bärnighausen TW, et al. Burden of neurological disorders across the us from 1990-2017: a global burden of disease study. JAMA Neurol 2021 Feb 01;78(2):165-176 [FREE Full text] [doi: 10.1001/jamaneurol.2020.4152] [Medline: 33136137]
21. Dopheide JA. Insomnia overview: epidemiology, pathophysiology, diagnosis and monitoring, and nonpharmacologic therapy. Am J Manag Care 2020 Mar 12;26(4 Suppl):S76-S84 [FREE Full text] [doi: 10.37765/ajmc.2020.42769] [Medline: 32282177]
22. Lagan S, Ramakrishnan A, Lamont E, Ramakrishnan A, Frye M, Torous J. Digital health developments and drawbacks: a review and analysis of top-returned apps for bipolar disorder. Int J Bipolar Disord 2020 Dec 01;8(1):39 [FREE Full text] [doi: 10.1186/s40345-020-00202-4] [Medline: 33259047]
23. Lagan S, Aquino P, Emerson MR, Fortuna K, Walker R, Torous J. Actionable health app evaluation: translating expert frameworks into objective metrics. NPJ Digit Med 2020 Jul 30;3(1):1-8. [doi: 10.1038/s41746-020-00312-4]
24. Tangari G, Ikram M, Ijaz K, Kaafar MA, Berkovsky S. Mobile health and privacy: cross sectional study. BMJ 2021 Jun 16;373:n1248 [FREE Full text] [doi: 10.1136/bmj.n1248] [Medline: 34135009]
25. Albert Henry T. FTC warns health apps to comply with health data-breach rules. AMA. 2021 Nov 29. URL: https://www.ama-assn.org/practice-management/digital/ftc-warns-health-apps-comply-health-data-breach-rules#:~:text=Health%20apps%20and%20connected%20devices,Federal%20Trade%20Commission%20(FTC) [accessed 2022-06-01]
26. Muñoz AO, Camacho E, Torous J. Marketplace and literature review of Spanish language mental health apps. Front Digit Health 2021 Feb 15;3:615366 [FREE Full text] [doi: 10.3389/fdgth.2021.615366] [Medline: 34713093]
27. 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]

Abbreviations

FTC: Federal Trade Commission
HIPAA: Health Insurance Portability and Accountability Act
MIND: Mobile Health Index and Navigation

Edited by L Buis; submitted 24.01.22; peer-reviewed by M Ainslie, M K., K Fadahunsi, G Noomen; comments to author 28.03.22; revised version received 17.05.22; accepted 25.05.22; published 21.06.22.

Please cite as:
Minen MT, George A, Camacho E, Yao L, Sahu A, Campbell M, Soviero M, Hossain Q, Verma D, Torous J
Assessment of Smartphone Apps for Common Neurologic Conditions (Headache, Insomnia, and Pain): Cross-sectional Study
JMIR Mhealth Uhealth 2022;10(6):e36761
URL: https://mhealth.jmir.org/2022/6/e36761
doi:10.2196/36761
PMID:35727625

Original Paper

Validation Parameters of Patient-Generated Data for Digitally Recorded Allergic Rhinitis Symptom and Medication Scores in the @IT.2020 Project: Exploratory Study

Stephanie Dramburg[1], MD; Serena Perna[1], MS; Marco Di Fraia[1], MD; Salvatore Tripodi[2,3], MD; Stefania Arasi[1,4], MD; Sveva Castelli[1]; Danilo Villalta[5], MD; Francesca Buzzulini[5], MD; Ifigenia Sfika[2], MD; Valeria Villella[2]; Ekaterina Potapova[1], MS; Maria Antonia Brighetti[6]; Alessandro Travaglini[6]; Pierluigi Verardo[7]; Simone Pelosi[8], Eng; Paolo Maria Matricardi[1], MD

[1]Department of Pediatric Respiratory Medicine, Immunology and Critical Care Medicine, Charité-Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

[2]Pediatric Allergology Unit, Sandro Pertini Hospital, Rome, Italy

[3]Allergology Service, Policlinico Casilino, Rome, Italy

[4]Translational Research in Pediatric Specialities Area, Division of Allergy, Bambino Gesù Children's Hospital, Scientific Institute for Research, Hospitalization and Healthcare, Rome, Italy

[5]Department of Immunology and Allergy, Santa Maria degli Angeli Hospital, Pordenone, Italy

[6]Department of Biology, University of Rome Tor Vergata, Rome, Italy

[7]Center of Aerobiology, Regional Agency for the Protection of the Environment, Pordenone, Italy

[8]TPS Production, Rome, Italy

Corresponding Author:
Paolo Maria Matricardi, MD
Department of Pediatric Respiratory Medicine, Immunology and Critical Care Medicine
Charité-Universitätsmedizin Berlin
Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin
Augustenburgerplatz, 1
Berlin, 13353
Germany
Phone: 49 30 450 566 406
Email: paolo.matricardi@charite.de

Abstract

Background: Mobile health technologies enable allergists to monitor disease trends by collecting daily patient-reported outcomes of allergic rhinitis. To this end, patients with allergies are usually required to enter their symptoms and medication repetitively over long time periods, which may present a risk to data completeness and quality in the case of insufficient effort reporting. Completeness of patient's recording is easily measured. In contrast, the intrinsic quality and accuracy of the data entered by the patients are more elusive.

Objective: The aim of this study was to explore the association of adherence to digital symptom recording with a predefined set of parameters of the patient-generated symptom and medication scores and to identify parameters that may serve as proxy measure of the quality and reliability of the information recorded by the patient.

Methods: The @IT.2020 project investigates the diagnostic synergy of mobile health and molecular allergology in patients with seasonal allergic rhinitis. In its pilot phase, 101 children with seasonal allergic rhinitis were recruited in Rome and instructed to record their symptoms, medication intake, and general conditions daily via a mobile app (AllergyMonitor) during the relevant pollen season. We measured adherence to daily recording as the percentage of days with data recording in the observation period. We examined the patient's trajectories of 3 disease indices (Rhinoconjunctivitis Total Symptom Score [RTSS], Combined Symptom and Medication Score [CSMS], and Visual Analogue Scale [VAS]) as putative proxies of data quality with the following 4 parameters: (1) intravariation index, (2) percentage of zero values, (3) coefficient of variation, and (4) percentage of changes in trend. Lastly, we examined the relationship between adherence to recording and each of the 4 proxy measures.

Results: Adherence to recording ranged from 20% (11/56) to 100% (56/56), with 64.4% (65/101) and 35.6% (36/101) of the patients' values above (highly adherent patients) or below (low adherent patients) the threshold of 80%, respectively. The percentage of zero values, the coefficient of variation, and the intravariation index did not significantly change with the adherence to recording. By contrast, the proportion of changes in trend was significantly higher among highly adherent patients, independently from the analyzed score (RTSS, CSMS, and VAS).

Conclusions: The percentage of changes in the trend of RTSS, CSMS, and VAS is a valuable candidate to validate the quality and accuracy of the data recorded by patients with allergic rhinitis during the pollen season. The performance of this parameter must be further investigated in real-life conditions before it can be recommended for routine use in apps and electronic diaries devoted to the management of patients with allergic rhinitis.

(JMIR Mhealth Uhealth 2022;10(6):e31491) doi:10.2196/31491

KEYWORDS

allergic rhinitis; symptom scores; patient-generated data; patient-reported outcomes; mHealth; mobile health; health applications; allergies; allergy monitor; digital health; medication scores

Introduction

Digital and mobile health technologies are creating new perspectives in many areas of research and medical care. One important aspect in both fields is the ability to easily collect patient-generated data via smartphone apps and connected devices such as wearables, diagnostic tools, or environmental sensors [1-5]. Although the use of patient-reported outcomes has become popular over the last decade [6-9], it is not sure how accurately the collected data represent the patient's state, as recording is done without supervision [10,11]. In particular, daily reporting over a longer time period may be perceived as challenging and cause a certain degree of reporting fatigue. The risk of potentially lower quality owing to disengaged survey respondents has been described with different terms, most recently in the field of psychology as "*insufficient effort reporting*" [11,12]. However, consensus on methodologies assessing the quantity and quality of entered data is still missing. Proposed methodologies include the (1) response pattern approach [13], (2) response time approach [14], (3) infrequency approach [15], (4) inconsistency approach [16], and (5) the number of unanswered questions. Most of these methodologies, however, refer to single points of data collection and extended questionnaires, which make their application in a setting with daily data recording via smartphone apps difficult to impossible. As digital methods of data collection via openly available mobile apps usually generate very large data sets, new challenges for the analysis and interpretation apply, such as the lack of standard measures [17]. The importance of unified approaches to data recording has recently been underlined in the context of patient adherence, and computational solution approaches were published to support uniform data formats [18]. A representative example for the daily acquisition of patient-generated data are mobile apps related to seasonal allergic rhinitis [19,20]. A variety of apps has been published for patients with pollen allergies, providing exposure forecasts, individualized symptom prediction, symptom and medication diaries, and in some cases, the possibility of exchanging recorded data with the attending physician [3,10,21-23]. Although several studies have shown the potential of mobile technologies for research purposes and clinical disease management, only few address the topic of data quality and validation [24,25]. The purpose of this study is to retrospectively investigate 4 putative validation criteria to assess the quality of data longitudinally collected by patients with seasonal allergic rhinitis and defined as follows: (1) intravariation index, (2) percentage of zero values, (3) coefficient of variation, and (4) percentage of changes in trends. To this end, we have taken advantage of the patient-generated symptom and medication data set, which has been acquired via a mobile app in a cohort of patients with seasonal allergic rhinitis in the context of the @IT.2020 pilot project [26,27].

Methods

Study Design

The @IT.2020 pilot project is an observational clinical study on the impact of component-resolved diagnosis and digital symptom recording on the diagnosis of pollen allergy. In the context of this project, 101 patients experiencing seasonal allergic rhinitis were recruited in the Sandro Pertini Hospital in Rome. The detailed study protocol has been published previously [26,27]. Briefly, recruited patients underwent a medical examination first (T0), including skin prick testing, blood sampling, and clinical questionnaires. At the end of the visits, patients were instructed on the use of the AllergyMonitor (TPS Software Production) mobile app to monitor their symptoms of the eyes, nose, and lungs, as well as medication intake and the impact of allergy symptoms on their daily activities during an individual study period. After the monitoring period, all patients underwent a second medical examination (T1), including a repetition of the initial clinical questionnaires focused on the past pollen season.

Ethics Approval

The study design and procedures were approved by the local ethics committee "Comitato Etico Indipendente Lazio 2" (study 10-16, Protocol 9871—01/02/2016).

AllergyMonitor App

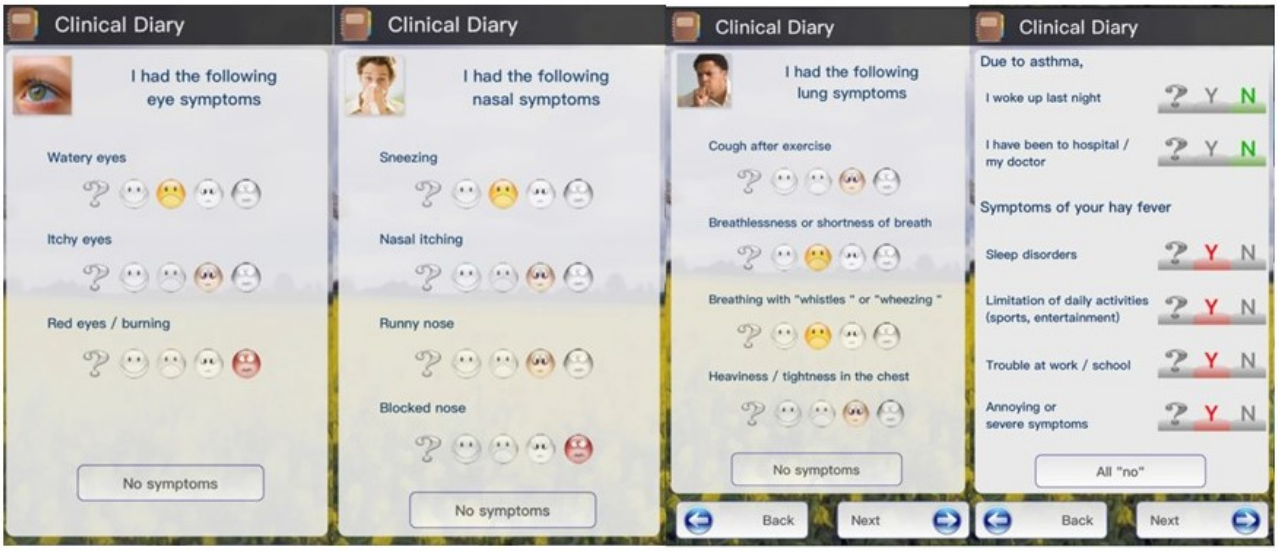
AllergyMonitor is a CE-certified smartphone app designed for the daily reporting of allergic symptoms of the eyes, nose, and lungs. Further, the impact of allergic symptoms on daily activities and sleep as well as the medication intake were recorded. In order to facilitate the correct recording of medication intake, the study doctor registered the patients' individual medication via the back end of the app, and the

patient's tailored drug name and administration schedule appeared in the app's front end.

Symptom and Medication Scores

The following symptom and symptom medication scores were used in this study: Rhinoconjunctivitis Total Symptom Score (RTSS, 0-18 points) [28], Combined Symptom and Medication Score (CSMS, 0-6 points) [29], and Visual Analogue Scale (VAS, 0-10 points) [30]. RTSS and CSMS were calculated automatically by the app for every reporting day on the basis of 4 questions on nasal symptoms (sneezing, rhinorrhea, nasal pruritus, nasal congestion), 2 questions on ocular symptoms (itchy eyes, watery eyes), and 3 questions on medication intake (antihistaminic drugs, local steroids, systemic steroids). The severity of each of the symptoms was also measured by the patient using 4 different emoticons (Figure 1), each one representing a distinct severity grade (no symptoms, mild, moderate, or severe). Overall, severity was also measured using VAS in response to the question "How do you feel in relation to your allergic symptoms today?"

Figure 1. Screenshots of the AllergyMonitor app, indicating the emoticons used to assess symptom severity.



Adherence to Electronic Diary Recording

Retrospective reporting of symptoms for missed days was only possible within 48 hours. After 2 days of missed reporting, an automated reminder appeared on the patient's phone. After 3 days of missed recording, the patient received an individual email or phone call from the study center to ensure that no technical problems had occurred. Adherence was measured as the percentage of days with completed electronic diary recording in the monitoring period [26]. Patients with adherence to recording above or below the arbitrary threshold of 80% were defined as high or low adherence, respectively.

Validation Parameters

Data retrospectively obtained were summarized as numbers (n) and frequencies (%) if they were categorical and as mean (SD) or median (IQR) if quantitative. The prevalence of atopic sensitization to airborne allergens was evaluated for every patient via skin prick test results (cutoff for positivity: a wheal size of ≥3 mm). For every pollen period considered, adherence values were calculated for each patient. Four tentative parameters for validation of the data quality were calculated as follows.

Intravariation Index

For each i^{th} subject ($i = 1,,,,n$), (1) the percentage of the number of variations between 2 consecutive days was calculated, , where $t= 1,,,,,,,$ T is the indicator of time point considered, y is the value of the symptom medication score considered, and I is a binary variable, which is 1 when the condition in the brackets is verified, 0 otherwise; (2) the individual variation range was calculated (rVAR = (max(y1, …, yT) – min(y1, …, yT))/S, where S represents the unit increase for each symptom medication score, that is, S=1 for VAS and RTSS and S=0.167 for CSMS); and (3) the average of all individual intravariation index values was calculated, , for each symptom score or symptom medication score.

Percentage of Zero Values

Average of individual percentages of zero values (100*number of zero compiled values/actual compiled days)

Coefficient of Variation

Average of individual coefficient of variation (100*SD/mean)

Percentage of Changes in Trend

Average of individual percentages of changes in severity trends (worsening=plus; stability=zero; improvement=minus) between 2 consecutive values of symptom medication score

Statistical Analysis

Spearman rank correlation coefficient was used to investigate the relationship between individual averages for each symptom medication score considered. The nonparametric Mann-Whitney *U* test was applied to compare the average values of quality indexes between 2 groups of subjects divided by their adherence of recording (<80% vs ≥80%). $P<.05$ was considered statistically significant. Statistical analyses were performed with R Core

Team (2018) version 3.5.2 (The R Project for Statistical Computing).

Results

Study Population and Pollen Season

This analysis consisted of 101 children (mean age 13.7 [SD 2.8] years) meeting the inclusion criteria for the @IT.2020 pilot study. Male gender was slightly more frequent (63/101, 62.4%), and the population was characterized by predominantly persistent allergic rhinitis symptoms by Allergic Rhinitis and its Impact on Asthma criteria, as assessed by retrospective questionnaire during T0. In addition to persistence, the severity of symptoms for 39.6% (40/101) of the patients was classified as moderate-to-severe (Table 1). The rate of patients with moderate-to-severe persistent symptoms increased to 70.3% (64/101) at the final study visit when being asked the same questionnaire concerning the past pollen season. The most frequent allergic comorbidities were oral allergy syndrome (32/101, 31.7%), atopic dermatitis (28/101, 27.7%), and allergic asthma (28/101, 27.7%). Most patients were sensitized to grass pollen, with 97% (98/101) having a positive skin prick test to Timothy grass and 90.1% (91/101) reacting to Bermuda grass (Table S1 of Multimedia Appendix 1). Grass pollen concentrations ranged from 0 to 199 grains/m^3 air. Season criteria of the European Academy of Allergy and Clinical Immunology [31] were adapted to the local setting, and the dates of whole grass pollen season from April 13 to July 28 as well as the peak grass pollen season between May 4 and June 28, 2016 are reported in this study (Figure S1 of Multimedia Appendix 1).

Table 1. Characteristics of the study population (N=101).

Characteristics	Value
Males, n (%)	63 (62.4)
Age (years), mean (SD)	13.7 (2.8)
Allergic rhinitis	
Age at onset (years), median (IQR)	6 (4-8)
Allergic rhinitis and its impact on asthma classification at first medical examination, n (%)	
Mild intermittent	19 (18.8)
Mild persistent	31 (30.7)
Moderate/severe intermittent	11 (10.9)
Moderate/severe persistent	40 (39.6)
Allergic rhinitis and its impact on asthma classification at second medical examination, n (%)	
Mild intermittent	6 (6.6)
Mild persistent	17 (18.7)
Moderate/severe intermittent	4 (4.4)
Moderate/severe persistent	64 (70.3)
Other allergic comorbidities, n (%)	
Allergic asthma	28 (27.7)
Oral Allergy Syndrome	32 (31.7)
Urticaria/angioedema	19 (19.2)
Atopic dermatitis	28 (27.7)
Gastrointestinal disorders	4 (4)
Anaphylaxis episode	10 (10.1)
Other	5 (5.1)

Adherence to Recording

During the grass pollen season (May 4 to June 28, 2016), 4003 single reports were collected, equaling an average adherence to recording of 70.8% (4003/5654). Over the period of 56 days, the individual number of filled questionnaires ranged from 11 (20%) to 56 (100%); 65 of the 101 patients (64.4%) were highly adherent to data collection. A delayed reporting start or an anticipated end [27] was observed, with 9 patients starting the monitoring 3 days or more after the start of the prescribed period and 12 patients ending the reporting ≥3 days before the prescribed end. Figure S2 of Multimedia Appendix 1 shows that 53 patients had a prescribed monitoring starting before the grass pollen season.

Interrelation Among RTSS, CSMS, and VAS

The RTSS and CSMS correlate well at a population level over time (Figures 2 and 3). Although these 2 scores are calculated

based on identical symptom questions with and without the integration of symptomatic medication, the VAS depicted in the bottom panel takes information from a separate question, filled with the same frequency and showing a similar trend over time. At the individual level, the average VAS score correlated well with both—the average symptom score (RTSS) and the average symptom-medication score (CSMS) (Figure 3A and B).

Figure 2. Average population values of Rhinoconjunctivitis Total Symptom Score (0-18 points), Combined Symptom and Medication Score (0-6 points) (both top panel), and Visual Analogue Scale (0-10 points) on impact of allergic symptoms on daily life (bottom panel) over time. CSMS: Combined Symptom and Medication Score; RTSS: Rhinoconjunctivitis Total Symptom Score; VAS: Visual Analogue Scale.

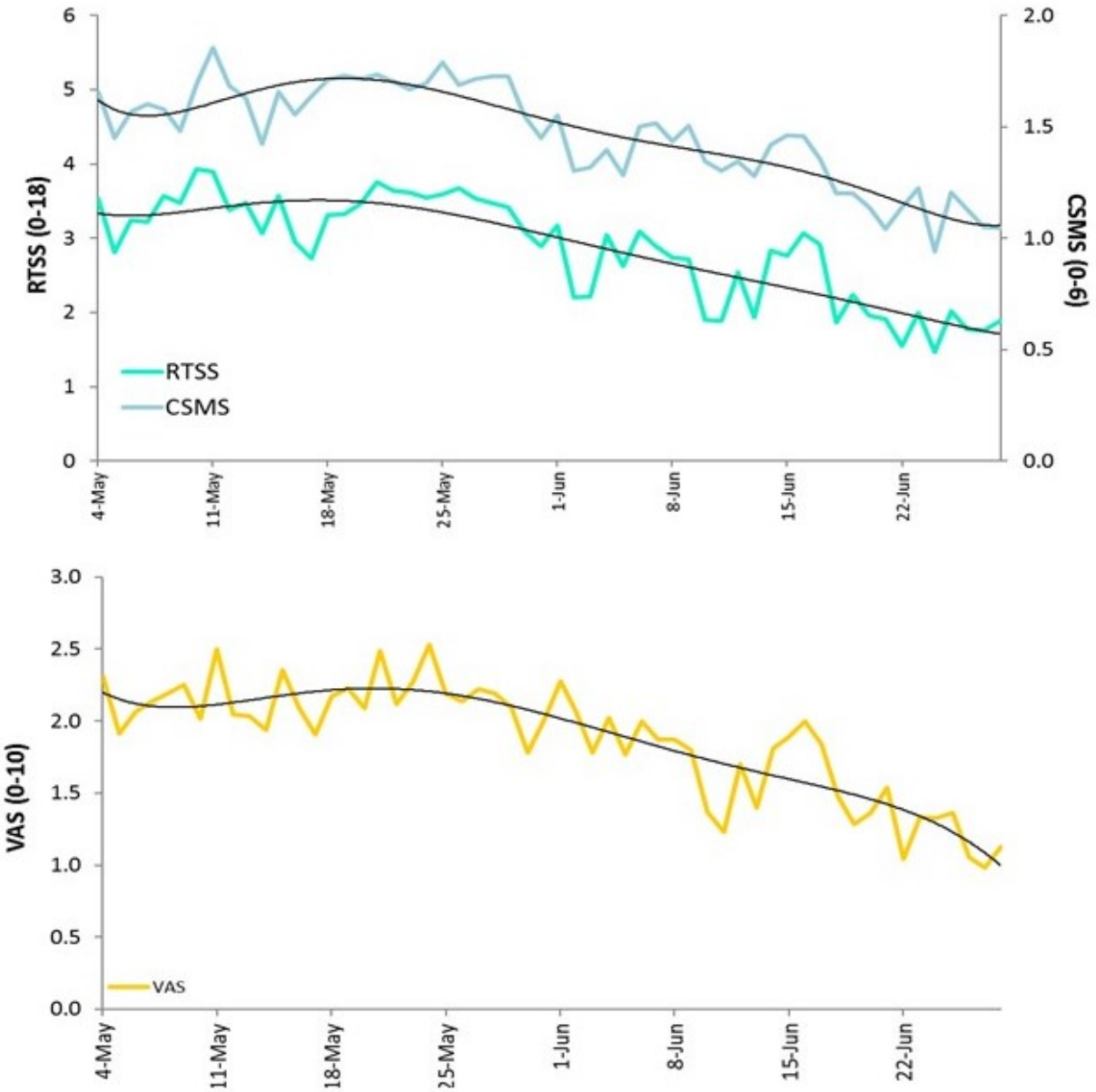
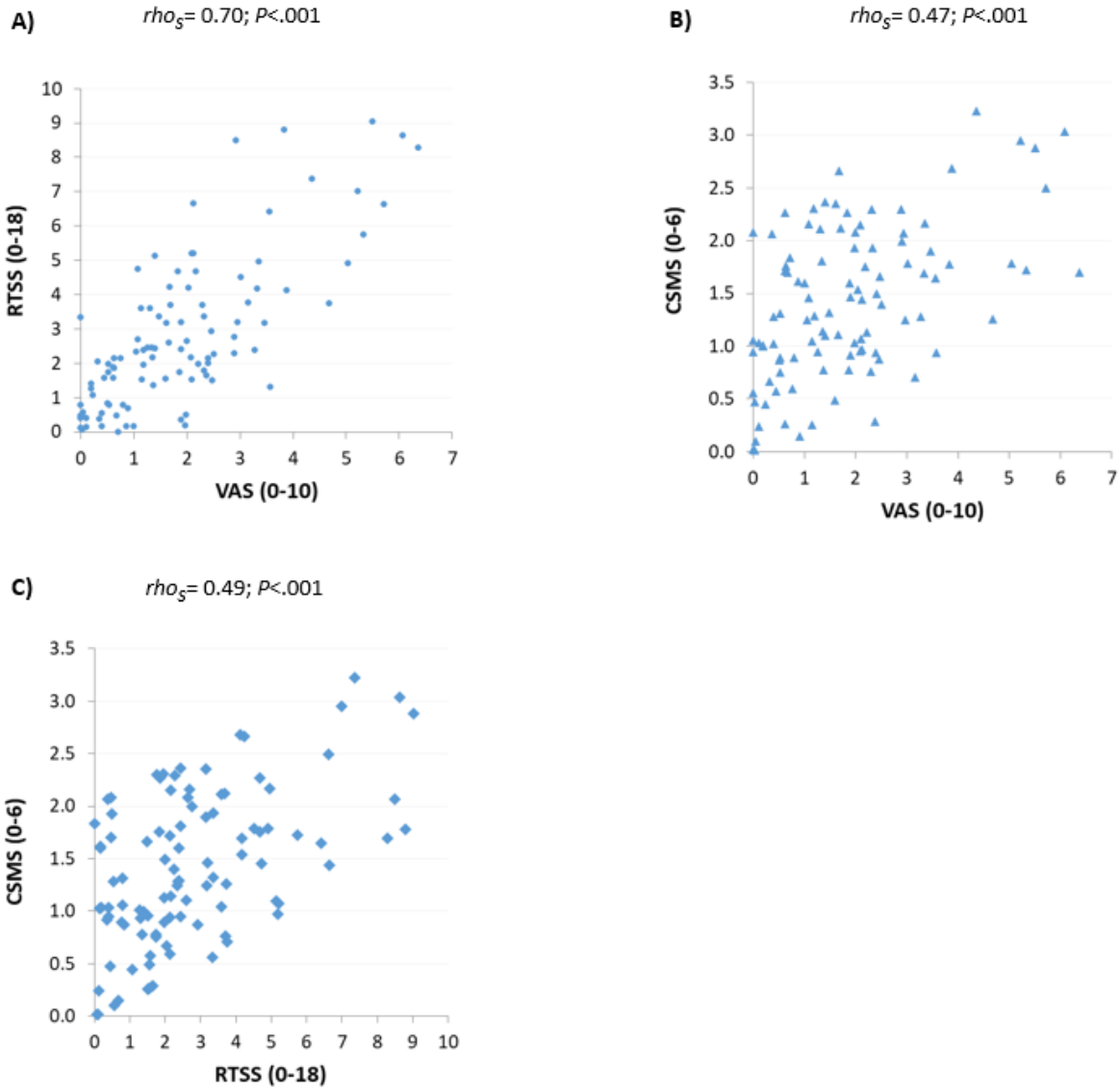


Figure 3. Correlation between individual averages of (A) Rhinoconjunctivitis Total Symptom Score versus Visual Analogue Scale, (B) Combined Symptom and Medication Score versus Visual Analogue Scale, and (C) Combined Symptom and Medication Score versus Rhinoconjunctivitis Total Symptom Score. CSMS: Combined Symptom and Medication Score; RTSS: Rhinoconjunctivitis Total Symptom Score; VAS: Visual Analogue Scale.



Quality Indices in RTSS, VAS, and CSMS

In order to assess the quality of recorded data, 4 different parameters were investigated within each of the 3 scores (RTSS, VAS, CSMS) (Table 2), and average values between highly and poorly adherent patients were compared (Table 3, Figure 4).

The highest diversity of data was observed in the VAS and the RTSS, as expressed by the intravariation index and the coefficient of variation. However, the CSMS was more homogeneous over time. As expected, the percentage of zero values was the lowest in the CSMS, whose average values were almost half of those observed in the RTSS and VAS. By contrast, the percentage of changes in the trend was quite similar and high for all the 3 scores, with values oscillating around 30% (Table 2).

Table 2. Quality of symptom and symptom-medication scores.

Quality index	Rhinoconjunctivitis Total Symptom Score, mean (95% CI)	Visual Analogue Scale, mean (95% CI)	Combined Symptom and Medication Score, mean (95% CI)
Intravariation index[a]	5.1 (4.5-5.6)	6.1 (5.5-6.7)	2.9 (2.6-3.2)
% of zero values[b]	38.9 (33.3-44.5)	43.3 (37.2-49.3)	20.4 (16.0-24.8)
Coefficient of variation[c]	134.8 (116.5-153.0)	138.1 (115.8-160.5)	86.3 (72.3-100.3)
% of changes in trends[d]	31.3 (28.2-34.4)	28.9 (25.6-32.3)	32.7 (29.6-35.7)

[a]Average of intravariation index by subjects; for each ith subject, (1) the percentage of the number of variations between 2 consecutive days is calculated,  , where *t*= 1,,,,,,, T is the indicator of time point considered, y is the value of the symptom medication score considered, and I is a binary variable, which is 1 when the condition in the brackets is verified, 0 otherwise; (2) the individual variation range has been calculated (rVAR = (max(y_1, …,y_T) – min(y1, …, y_T)) + 1/S, where S represents the unit increase for each symptom medication score, that is, S=1 for Visual Analogue Scale and Rhinoconjunctivitis Total Symptom Score and S=0.167 for Combined Symptom and Medication Score); and (3) the average of all individual intravariation index values was calculated,  , for each symptom score or symptom medication score.

[b]100*number of zero compiled values/actual compiled days.

[c]Average of individual coefficient of variation (100*SD/mean).

[d]Number of changes in trends (plus/minus/stable) within ith differences between 2 consecutive values of symptom medication score.

Table 3. Association between adherence and quality indexes.

	Rhinoconjunctivitis Total Symptom Score			Visual Analogue Scale			Combined Symptom and Medication Score		
	Adh[a]<80% (n=36), median (IQR)	Adh≥80% (n=65), median (IQR)	*P* value[b]	Adh<80% (n=36), median (IQR)	Adh≥80% (n=65), median (IQR)	*P* value	Adh<80% (n=36), median (IQR)	Adh≥80% (n=65), median (IQR)	*P* value
Intravariation index[c]	4 (3-6)	5 (4-6)	.08	5 (4-8)	7 (4-8)	.18	3 (2-3)	3 (2-4)	.32
% of zero values[d]	27 (12-72)	38 (17-55)	.66	22 (16-60)	43 (22-65)	.31	13 (0-31)	16 (4-28)	.61
Coefficient of variation[e]	105 (79-197)	99 (79-131)	.76	84 (60-115)	115 (76-158)	.06	77 (53-95)	74 (48-101)	.67
% of changes in trend[f]	25 (11-30)	34 (26-51)	<.001	23 (11-31)	33 (16-45)	.004	26 (18-30)	36 (28-51)	<.001

[a]Adh: adherence.

[b]Mann-Whitney *U* test was used to compare means among the 2 groups.

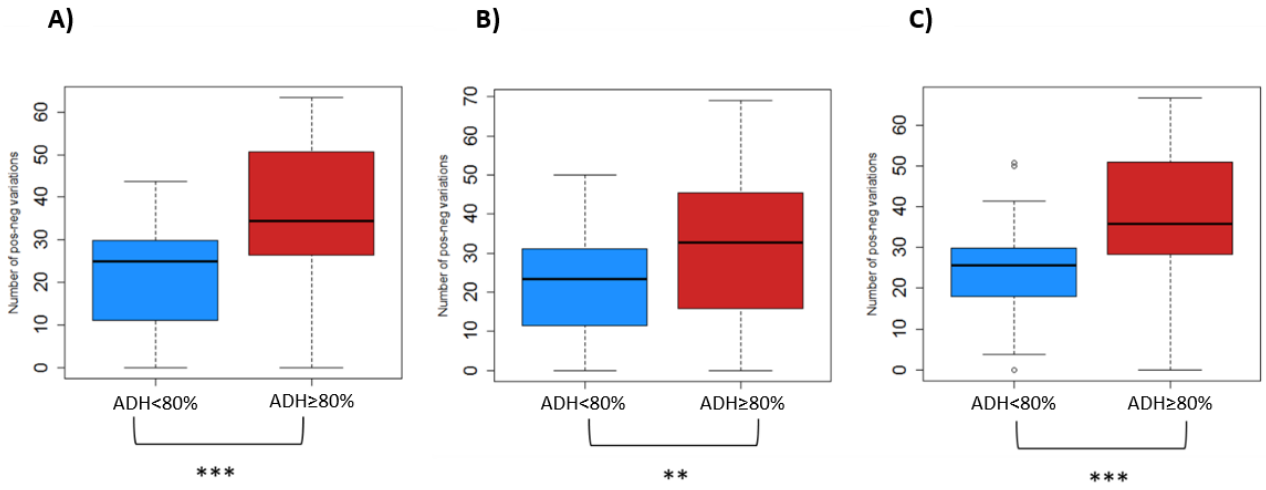
[c]Average of intravariation index by subjects; for each ith subject, (1) the number of variations between 2 consecutive days has been calculated,  , where *t*= 1,,,,,,, T is the indicator of the time point considered, y is the value of the symptom medication score considered, and I is a dummy variable, which is 1 when the condition in the brackets is verified, 0 otherwise; (2) the individual variation range has been calculated (rVAR = (max(y_1, …,y_T) – min(y1, …, y_T)) + 1/S, where S represents the unit increase for each symptom medication score, that is, S=1 for Visual Analogue Scale and Rhinoconjunctivitis Total Symptom Score and S=0.167 for Combined Symptom and Medication Score); and (3) the average of all individual intravariation index values was calculated,  , for each symptom score or symptom medication score.

[d]100*number of zero compiled values/actual compiled days.

[e]Average of individual coefficient of variation (100*SD/mean).

[f]Number of changes in trends (plus/minus/stable) within the differences between 2 consecutive values of symptom medication score.

Figure 4. Changes in trends (positive/negative changes) of recorded (A) Rhinoconjunctivitis Total Symptom Score, (B) Visual Analogue Scale, and (C) Combined Symptom and Medication Score values among patients with adherence to recording of <80% versus ≥80%. ADH: adherence. ***P*<.01; ****P*<.001.



Relationship of Quality Index Values With Adherence to Recording

No significant differences were observed between the groups of highly and poorly adherent patients with regard to the average intravariation index, percentage of zero values, and the coefficient of variation. Of note, the coefficient of variation of the VAS score was higher among highly than poorly adherent patients, but the difference was only marginally significant (*P*=.06). Similarly, the average intravariation index of the RTSS was higher among patients with adherence to recording of 80% or more, but the difference did not reach statistical significance (*P*=.08). In contrast, the percentage of changes in trend was significantly lower among patients who recorded their symptoms on less than 80% of the prescribed days compared to highly adherent patients, considering each of the 3 indexes: RTSS (*P*<.001), VAS (*P*<.005), and CSMS (*P*<.001) (Table 3, Figure 3).

Discussion

We retrospectively analyzed patient-generated data recorded during the grass pollen season by patients with seasonal allergic rhinitis to grasses. The collected data contained information on daily symptoms (RTSS), a combination of symptoms and medication intake (CSMS), and the overall impact of pollen allergy on daily life (VAS). Our analysis shows that (1) RTSS, CSMS, and VAS trajectories correlate well over time at population level; (2) VAS average values correlate well with average values of RTSS and CSMS at individual level; (3) the percentage of days with a change in trend during the observation period is higher in patients with high adherence to recording; and (4) other investigated parameters such as the percentage of zero values, the coefficient of variation, and the intravariation index are not significantly different among highly versus poorly adherent patients. Overall, our results suggest that the percentage of days with a change in trend deserves further investigation in a prospective study as a proxy of data quality in patients monitoring their pollen allergy with an electronic diary app.

Electronic diaries are increasingly produced and used in medicine, particularly in allergology. Nevertheless, studies focusing on the accuracy and completeness of the patient-generated information collected via electronic diaries are substantially missing. This is of high priority, as data validation is a prerequisite of any scientific or clinical use of the information collected through mobile apps from patients. A recent study demonstrated that daily monitoring with a VAS score has a high intrarater reliability and medium-high validity, reliability, and responsiveness, suggesting the validity of this simple methodology in monitoring disease impact on the patient's daily life [32]. Along the same line of evidence, our study demonstrates that VAS correlates well with complex measurements such as RTSS and CSMS, both in terms of trajectory at the population level (Figure 1) and as average values at the individual level (Figure 2).

The percentage of days with changes in trend within the registration period is an interesting parameter, as it can be examined within the context of any trajectory, independently from the structure of the algorithm generating the clinical score. Therefore, this parameter can be applied to VAS, RTSS, CSMS, or any other index that will be generated and validated in the future. We speculate that patients whose personal interest in the recording of their electronic diary is lower, may still be adherent but inaccurate, replicating the same pattern of values every day. The day-to-day variability of pollen counts coupled with daily variability of exposure to pollen as well as the use of preventative medication may impact markedly on day-to-day variability in symptom score. A patient highly adherent to recording is better placed to record the symptom variability and thus more likely to report a higher number of changes in trend. With regard to the use of electronic diaries in clinical practice and research, a tool to predict the quantity and quality of expected data collection would be helpful. Unfortunately, to our knowledge, no such tools exist at the moment. In a previous approach to assess and predict the adherence of patients to symptom recording of patients with pollen allergies, we observed an association between the reporting behavior between

the 7th and 21st day of recording compared to the rest of the monitoring period (up to 70 days) [27].

The lack of a previously established methodology also justifies our explorative approach in the attempt to identify new statistical methods or methods generated in other contexts to address a novel research question. We therefore speculate that future studies will adopt similar approaches and generate new and even more precise methodologies to answer the same research questions. Further, an expansion to other chronic diseases for which digital data collection has been well adopted, for example, asthma, will be of great value.

We are aware of the limitations of our analysis. First of all, we retrospectively examined in an opportunistic approach a database already generated with different targets. Consequently, our paper can only generate the hypothesis that the percentage of changes in trend is a valuable parameter measuring the quality of patient-recorded data. Unfortunately, other important parameters such as clinical validity could not be investigated within this data set, as this parameter should be investigated independently from the adherence to reporting. Therefore, before any use of changes in trend as a parameter in clinical practice, our hypothesis must be prospectively proven in studies designed with this specific scope. Second, we limited our investigation period to a maximum of 56 days of recording and in the context of a clinical investigation. The generalizability of our conclusions to a real-life setting and to longer periods of monitoring are also to be proven in real-world studies and longer observation periods. Third, our study population was composed of children, whose electronic diary recording is partially (in general until the age of 14 years) performed with the assistance of parents and whose influence on the reliability of data should also be accounted for. Fourth, we have used adherence to recording as a reference parameter under the assumption that patients more compliant in regularly filling their electronic diaries are also those whose data are more reliable. This assumption also should be proven in a new prospective study by adopting external quality standards not affected by recording patterns. Fifth, there is no possibility of correlating medication usage with adherence to digital symptom recording.

In conclusion, our retrospective analysis identifies the percentage of changes in trend in the trajectory of RTSS, CSMS, and VAS as a parameter, intrinsic to the trajectory itself, thereby representing a valuable candidate as proxy measure of data quality. This hypothesis deserves now to be investigated in prospective studies.

Acknowledgments

This study was supported by an unrestricted educational grant from Euroimmun (code 118583). We thank all participants and the entire study team. SA was supported by the European Academy of Allergy and Clinical Immunology Fellowship Award. This study has been supported by an unrestricted grant from Euroimmun. The Informatics Platform AllergyCARD and the AllergyMonitor app have been kindly provided by TPS Software Production.

Conflicts of Interest

PMM reports grants and personal fees from TPS Software Production, outside the submitted work. ST is a cofounder of TPS Software Production, which has developed the AllergyMonitor app. S Pelosi is a cofounder of TPS Software Production. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Supplementary data.
[DOCX File , 266 KB - mhealth_v10i6e31491_app1.docx]

References

1. Nahum JL, Fu MR, Scagliola J, Rodorigo M, Tobik S, Guth A, et al. Real-time electronic patient evaluation of lymphedema symptoms, referral, and satisfaction: a cross-sectional study. Mhealth 2021;7:20 [FREE Full text] [doi: 10.21037/mhealth-20-118] [Medline: 33898589]
2. Gomis-Pastor M, Mirabet Perez S, Roig Minguell E, Brossa Loidi V, Lopez Lopez L, Ros Abarca S, et al. Mobile Health to Improve Adherence and Patient Experience in Heart Transplantation Recipients: The mHeart Trial. Healthcare (Basel) 2021 Apr 14;9(4):463 [FREE Full text] [doi: 10.3390/healthcare9040463] [Medline: 33919899]
3. Tripodi S, Giannone A, Sfika I, Pelosi S, Dramburg S, Bianchi A, et al. Digital technologies for an improved management of respiratory allergic diseases: 10 years of clinical studies using an online platform for patients and physicians. Ital J Pediatr 2020 Jul 25;46(1):105 [FREE Full text] [doi: 10.1186/s13052-020-00870-z] [Medline: 32711557]
4. Khusial RJ, Honkoop PJ, Usmani O, Soares M, Simpson A, Biddiscombe M, myAirCoach study group. Effectiveness of myAirCoach: A mHealth Self-Management System in Asthma. J Allergy Clin Immunol Pract 2020 Jun;8(6):1972-1979.e8 [FREE Full text] [doi: 10.1016/j.jaip.2020.02.018] [Medline: 32142961]
5. Steel PAD, Bodnar D, Bonito M, Torres-Lavoro J, Eid DB, Jacobowitz A, et al. MyEDCare: Evaluation of a Smartphone-Based Emergency Department Discharge Process. Appl Clin Inform 2021 Mar;12(2):362-371. [doi: 10.1055/s-0041-1729165] [Medline: 33910262]

6. Richter JG, Nannen C, Chehab G, Acar H, Becker A, Willers R, et al. Mobile App-based documentation of patient-reported outcomes - 3-months results from a proof-of-concept study on modern rheumatology patient management. Arthritis Res Ther 2021 Apr 19;23(1):121 [FREE Full text] [doi: 10.1186/s13075-021-02500-3] [Medline: 33874994]
7. Benson T. Person-specific outcome measure (PSO) for use in primary and community care. BMJ Open Qual 2021 Apr;10(2):1767-1768 [FREE Full text] [doi: 10.1136/bmjoq-2021-001379] [Medline: 33931439]
8. Avery P. Using e-health tools and PROMs to support self-management in patients with inflammatory bowel disease. Br J Nurs 2021 Apr 08;30(7):394-402 [FREE Full text] [doi: 10.12968/bjon.2021.30.7.394] [Medline: 33830794]
9. Austin E, Lee JR, Amtmann D, Bloch R, Lawrence SO, McCall D, et al. Use of patient-generated health data across healthcare settings: implications for health systems. JAMIA Open 2020 Apr;3(1):70-76 [FREE Full text] [doi: 10.1093/jamiaopen/ooz065] [Medline: 32607489]
10. Matricardi PM, Dramburg S, Alvarez-Perea A, Antolín-Amérigo D, Apfelbacher C, Atanaskovic-Markovic M, Cavkaytar, et al. The role of mobile health technologies in allergy care: An EAACI position paper. Allergy 2020 Feb;75(2):259-272. [doi: 10.1111/all.13953] [Medline: 31230373]
11. Bousquet J, Ansotegui IJ, Anto JM, Arnavielhe S, Bachert C, Basagaña X, et al. Mobile Technology in Allergic Rhinitis: Evolution in Management or Revolution in Health and Care? J Allergy Clin Immunol Pract 2019;7(8):2511-2523. [doi: 10.1016/j.jaip.2019.07.044] [Medline: 31445223]
12. Huang J, DeSimone J. Insufficient Effort Responding as a Potential Confound between Survey Measures and Objective Tests. J Bus Psychol 2020 Aug 22;36(5):807-828 [FREE Full text] [doi: 10.1007/s10869-020-09707-2]
13. Baer RA, Kroll LS, Rinaldo J, Ballenger J. Detecting and Discriminating Between Random Responding and Overreporting on the MMPI-A. Journal of Personality Assessment 1999 Apr;72(2):308-320. [doi: 10.1207/s15327752jp720213]
14. Piedmont RL, McCrae RR, Riemann R, Angleitner A. On the invalidity of validity scales: evidence from self-reports and observer ratings in volunteer samples. J Pers Soc Psychol 2000 Mar;78(3):582-593. [doi: 10.1037//0022-3514.78.3.582] [Medline: 10743882]
15. Costa P, McCrae R. The revised NEO personality inventory (NEO-PI-R). In: Boyle GJ, Matthews G, Saklofske DH, editors. The SAGE Handbook of Personality Theory and Assessment (Volume 2): Personality Measurement and Testing. California: SAGE Publications Ltd; 2008.
16. Wise SL, DeMars CE. An Application of Item Response Time: The Effort-Moderated IRT Model. J Educational Measurement 2006 Mar;43(1):19-38. [doi: 10.1111/j.1745-3984.2006.00002.x]
17. Kardas P, Aguilar-Palacio I, Almada M, Cahir C, Costa E, Giardini A, et al. The Need to Develop Standard Measures of Patient Adherence for Big Data: Viewpoint. J Med Internet Res 2020 Aug 27;22(8):e18150 [FREE Full text] [doi: 10.2196/18150] [Medline: 32663138]
18. Dima AL, Dediu D. Computation of adherence to medication and visualization of medication histories in R with AdhereR: Towards transparent and reproducible use of electronic healthcare data. PLoS One 2017;12(4):e0174426 [FREE Full text] [doi: 10.1371/journal.pone.0174426] [Medline: 28445530]
19. Florack J, Brighetti MA, Perna S, Pizzulli A, Pizzulli A, Tripodi S, et al. Comparison of six disease severity scores for allergic rhinitis against pollen counts a prospective analysis at population and individual level. Pediatr Allergy Immunol 2016 Jun;27(4):382-390. [doi: 10.1111/pai.12562] [Medline: 26992008]
20. Bastl K, Kmenta M, Pessi A, Prank M, Saarto A, Sofiev M, et al. First comparison of symptom data with allergen content (Bet v 1 and Phl p 5 measurements) and pollen data from four European regions during 2009-2011. Sci Total Environ 2016 Apr 01;548-549:229-235. [doi: 10.1016/j.scitotenv.2016.01.014] [Medline: 26802351]
21. Bousquet J, Caimmi DP, Bedbrook A, Bewick M, Hellings PW, Devillier P, et al. Pilot study of mobile phone technology in allergic rhinitis in European countries: the MASK-rhinitis study. Allergy 2017 Jun;72(6):857-865. [doi: 10.1111/all.13125] [Medline: 28072463]
22. Bédard A, Basagaña X, Anto JM, Garcia-Aymerich J, Devillier P, Arnavielhe S, MASK study group. Mobile technology offers novel insights into the control and treatment of allergic rhinitis: The MASK study. J Allergy Clin Immunol 2019 Jul;144(1):135-143.e6. [doi: 10.1016/j.jaci.2019.01.053] [Medline: 30951790]
23. Bastl K, Bastl M, Bergmann K, Berger M, Berger U. Translating the Burden of Pollen Allergy Into Numbers Using Electronically Generated Symptom Data From the Patient's Hayfever Diary in Austria and Germany: 10-Year Observational Study. J Med Internet Res 2020 Feb 21;22(2):e16767 [FREE Full text] [doi: 10.2196/16767] [Medline: 32130130]
24. Caimmi D, Baiz N, Tanno LK, Demoly P, Arnavielhe S, Murray R, MASK Study Group. Validation of the MASK-rhinitis visual analogue scale on smartphone screens to assess allergic rhinitis control. Clin Exp Allergy 2017 Dec;47(12):1526-1533. [doi: 10.1111/cea.13025] [Medline: 28886234]
25. Silver JD, Spriggs K, Haberle SG, Katelaris CH, Newbigin EJ, Lampugnani ER. Using crowd-sourced allergic rhinitis symptom data to improve grass pollen forecasts and predict individual symptoms. Sci Total Environ 2020 Jun 10;720:137351. [doi: 10.1016/j.scitotenv.2020.137351] [Medline: 32325552]
26. Arasi S, Castelli S, Di Fraia M, Villalta D, Tripodi S, Perna S, et al. @IT2020: An innovative algorithm for allergen immunotherapy prescription in seasonal allergic rhinitis. Clin Exp Allergy 2021 Jun;51(6):821-828. [doi: 10.1111/cea.13867] [Medline: 33759270]

27. Di Fraia M, Tripodi S, Arasi S, Dramburg S, Castelli S, Villalta D, et al. Adherence to Prescribed E-Diary Recording by Patients With Seasonal Allergic Rhinitis: Observational Study. J Med Internet Res 2020 Mar 16;22(3):e16642 [FREE Full text] [doi: 10.2196/16642] [Medline: 32175909]
28. Canonica GW, Baena-Cagnani CE, Bousquet J, Bousquet PJ, Lockey RF, Malling H, et al. Recommendations for standardization of clinical trials with Allergen Specific Immunotherapy for respiratory allergy. A statement of a World Allergy Organization (WAO) taskforce. Allergy 2007 Mar;62(3):317-324 [FREE Full text] [doi: 10.1111/j.1398-9995.2006.01312.x] [Medline: 17298350]
29. Pfaar O, Demoly P, Gerth van Wijk R, Bonini S, Bousquet J, Canonica GW, European Academy of AllergyClinical Immunology. Recommendations for the standardization of clinical outcomes used in allergen immunotherapy trials for allergic rhinoconjunctivitis: an EAACI Position Paper. Allergy 2014 Jul;69(7):854-867. [doi: 10.1111/all.12383] [Medline: 24761804]
30. Bousquet PJ, Combescure C, Neukirch F, Klossek JM, Méchin H, Daures J, et al. Visual analog scales can assess the severity of rhinitis graded according to ARIA guidelines. Allergy 2007 Apr;62(4):367-372. [doi: 10.1111/j.1398-9995.2006.01276.x] [Medline: 17362246]
31. Pfaar O, Bastl K, Berger U, Buters J, Calderon MA, Clot B, et al. Defining pollen exposure times for clinical trials of allergen immunotherapy for pollen-induced rhinoconjunctivitis - an EAACI position paper. Allergy 2017 May;72(5):713-722. [doi: 10.1111/all.13092] [Medline: 27874202]
32. Sousa-Pinto B, Eklund P, Pfaar O, Klimek L, Zuberbier T, Czarlewski W, et al. Validity, reliability, and responsiveness of daily monitoring visual analog scales in MASK-air®. Clin Transl Allergy 2021 Aug;11(7):e12062 [FREE Full text] [doi: 10.1002/clt2.12062] [Medline: 34567526]

Abbreviations

CSMS: Combined Symptom and Medication Score
RTSS: Rhinoconjunctivitis Total Symptom Score
VAS: Visual Analogue Scale

Edited by L Buis; submitted 23.06.21; peer-reviewed by D Antolín-Amérigo, D Ryan; comments to author 08.07.21; revised version received 30.08.21; accepted 07.12.21; published 03.06.22.

Please cite as:
Dramburg S, Perna S, Di Fraia M, Tripodi S, Arasi S, Castelli S, Villalta D, Buzzulini F, Sfika I, Villella V, Potapova E, Brighetti MA, Travaglini A, Verardo P, Pelosi S, Matricardi PM
Validation Parameters of Patient-Generated Data for Digitally Recorded Allergic Rhinitis Symptom and Medication Scores in the @IT.2020 Project: Exploratory Study
JMIR Mhealth Uhealth 2022;10(6):e31491
URL: https://mhealth.jmir.org/2022/6/e31491
doi:10.2196/31491
PMID:35657659

Original Paper

Trends in Heart Rate and Heart Rate Variability During Pregnancy and the 3-Month Postpartum Period: Continuous Monitoring in a Free-living Context

Fatemeh Sarhaddi[1], MSc; Iman Azimi[1], PhD; Anna Axelin[2,3,4], PhD; Hannakaisa Niela-Vilen[2], PhD; Pasi Liljeberg[1], PhD; Amir M Rahmani[5,6,7], MBA, PhD

[1]Department of Computing, University of Turku, Turku, Finland

[2]Department of Nursing Science, University of Turku, Turku, Finland

[3]Department of Obstetrics and Gynaecology, Turku University Hospital, Turku, Finland

[4]Faculty of Medicine, University of Turku, Turku, Finland

[5]Department of Computer Science, University of California, Irvine, Irvine, CA, United States

[6]School of Nursing, University of California, Irvine, Irvine, CA, United States

[7]Institute for Future Health, University of California, Irvine, Irvine, CA, United States

Corresponding Author:
Amir M Rahmani, MBA, PhD
Department of Computer Science
University of California, Irvine
Donald Bren Hall, 6210
Irvine, CA, 92697
United States
Phone: 1 949 824 3590
Email: a.rahmani@uci.edu

Abstract

Background: Heart rate variability (HRV) is a noninvasive method that reflects the regulation of the autonomic nervous system. Altered HRV is associated with adverse mental or physical health complications. The autonomic nervous system also has a central role in physiological adaption during pregnancy, causing normal changes in HRV.

Objective: The aim of this study was to assess trends in heart rate (HR) and HRV parameters as a noninvasive method for remote maternal health monitoring during pregnancy and 3-month postpartum period.

Methods: A total of 58 pregnant women were monitored using an Internet of Things–based remote monitoring system during pregnancy and 3-month postpartum period. Pregnant women were asked to continuously wear Gear Sport smartwatch to monitor their HR and HRV extracted from photoplethysmogram (PPG) signals. In addition, a cross-platform mobile app was used to collect background and delivery-related information. We analyzed PPG signals collected during the night and discarded unreliable signals by applying a PPG quality assessment method to the collected signals. HR, HRV, and normalized HRV parameters were extracted from reliable signals. The normalization removed the effect of HR changes on HRV trends. Finally, we used hierarchical linear mixed models to analyze the trends of HR, HRV, and normalized HRV parameters.

Results: HR increased significantly during the second trimester (P<.001) and decreased significantly during the third trimester (P=.006). Time-domain HRV parameters, average normal interbeat intervals (IBIs; average normal IBIs [AVNN]), SD of normal IBIs (SDNN), root mean square of the successive difference of normal IBIs (RMSSD), normalized SDNN, and normalized RMSSD decreased significantly during the second trimester (P<.001). Then, AVNN, SDNN, RMSSD, and normalized SDNN increased significantly during the third trimester (with P=.002, P<.001, P<.001, and P<.001, respectively). Some of the frequency-domain parameters, low-frequency power (LF), high-frequency power (HF), and normalized HF, decreased significantly during the second trimester (with P<.001, P<.001, and P=.003, respectively), and HF increased significantly during the third trimester (P=.007). In the postpartum period, normalized RMSSD decreased (P=.01), and the LF to HF ratio (LF/HF) increased significantly (P=.004).

Conclusions: Our study indicates the physiological changes during pregnancy and the postpartum period. We showed that HR increased and HRV parameters decreased as pregnancy proceeded, and the values returned to normal after delivery. Moreover,

our results show that HR started to decrease, whereas time-domain HRV parameters and HF started to increase during the third trimester. The results also indicated that age was significantly associated with HRV parameters during pregnancy and postpartum period, whereas education level was associated with HRV parameters during the third trimester. In addition, our results demonstrate the possibility of continuous HRV monitoring in everyday life settings.

(JMIR Mhealth Uhealth 2022;10(6):e33458) doi:10.2196/33458

KEYWORDS

heart rate; heart rate variability; pregnancy; postpartum; continuous monitoring; PPG; mobile phone

Introduction

Background

Heart rate variability (HRV) reflects alterations in the regulation of the autonomic nervous system. Substantial changes in autonomic nervous system, by implication in HRV, occur during pregnancy. Such physiological changes help to ensure the healthy development of the fetus [1]. Heart rate (HR) increases during pregnancy [2], whereas HRV parameters decrease; however, the values usually return to normal within a few months of the postpartum period [3-5].

In addition to physiological causes, changes in HRV during pregnancy may also reflect other issues; for example, certain physical or mental complications. Previous studies have shown that HRV during pregnancy may indicate hypertensive disorders [6,7] or pre-eclampsia [8,9]. Pregnant women with gestational hypertension have higher low-frequency power (LF) to high-frequency power (HF) ratio (LF/HF) in early pregnancy than those with normal pregnancies [6]. Regarding pre-eclampsia, women have lower HF than those with normal pregnancies, resulting in an increase in LF/HF in pre-eclamptic pregnancies [8,9]. Furthermore, HRV may reflect the state of mental health in pregnant women; the effects of depression [10] and anxiety [11,12] on HRV parameters during pregnancy have been studied. Pregnant women with depression have low 24-hour time-domain parameters [10], and anxiety during pregnancy has been shown to decrease HF and very low–frequency power [11]. HRV parameters may also illustrate the level of stress experienced by pregnant women [13,14]. Induced stress has been shown to decrease HF in pregnant women. Symptoms of anxiety may further be associated with stress, as pregnant women with anxiety had dampened stress reactivity [12]. In addition, the decrease in root mean square of the successive difference of normal interbeat intervals (IBIs; RMSSD) and HF was significantly less in mindful pregnant women who have better resources to cope with stress during pregnancy [15]. All pregnancy-related complications are important to be detected early in maternity care, to enable appropriate interventions to secure the health of the pregnant woman and her fetus. However, interpretation of HRV is demanding owing to the complexity of the human cardiac system; changes in and the behavior of HRV varies across individuals.

HRV parameters are usually measured using electrocardiogram (ECG) or photoplethysmogram (PPG). Electrocardiography is a noninvasive method for monitoring the electrical activity of the cardiovascular system using electrodes attached to the skin. It is the gold standard for monitoring HR and HRV parameters, but cannot be used for long-term monitoring. In contrast, photoplethysmography is an optical method for monitoring heart activity and is more convenient for use in home and free-life settings. It is an easy-to-implement method that is used in many clinical and commercial wearable devices. Therefore, it is increasingly used in remote health monitoring systems.

Most studies have investigated changes in HRV in an episodic manner, using 1-time ECG recording of pregnant women at different gestational weeks or during labor [16-20]. In addition, in most longitudinal studies, 10 to 30 minutes of ECG were recorded from pregnant women once per trimester or monthly during pregnancy [7,21-24] and postpartum period [4]. The recordings were performed while the women were resting in a predefined position (usually supine position) in a laboratory setting. Stein et al [1] conducted 24-hour HRV recordings with pregnant women 4 times during pregnancy and once before pregnancy. Continuous measurements of HRV during pregnancy and early postpartum period may provide new and valuable information about the HRV patterns.

Although existing studies have characterized changes in HRV during pregnancy and the postpartum period, they have been limited to short-time recordings of ECG signals a few times during pregnancy. Some of these studies compared pregnant women at different gestational weeks with nonpregnant women to identify HRV trends. However, comparing HRV from different individuals can be inaccurate because HRV is unique for each person and is dependent on various parameters such as age and sex among many other factors [25]. In addition, other studies have collected few ECG recordings from the same individuals. Thus, owing to the limited number of measurements, the results cannot reliably capture the changes. Moreover, only Stein et al [1] collected data in home-based settings, whereas all the other studies used laboratory settings to collect HRV parameters.

Objectives

In this paper, we aimed to analyze the nighttime HRV trends during pregnancy and postpartum period. To the best of our knowledge, this study is the first to collect continuous PPG signals from pregnant women, in everyday settings over a long period. We used an Internet of Things (IoT)-based system to collect PPG signals from 58 women, several times a day during pregnancy and the first 3-month postpartum period. The continuous monitoring of HRV parameters enabled us to accurately detect HRV trends regarding in-person and between-person differences. Moreover, we analyzed the trends of normalized HRV parameters. The normalization was performed based on average HR to remove the effect of HR changes on HRV parameters. In addition, we added age, BMI,

and education level to our analysis as controlling factors and analyzed their effects on HRV trends. In summary, the contributions of this study were as follows:

1. Continuous monitoring of HRV in pregnant women, in everyday settings using a customized, remote, IoT-based monitoring system.
2. Analyzing HRV trends during pregnancy and postpartum period during the night.
3. Analyzing normalized HRV trends during pregnancy and postpartum period during the night.

Methods

Study Design

HRV parameter trends during pregnancy and postpartum period were investigated in a longitudinal study using an IoT-based system. The system used a smartwatch to remotely collect HRV parameters and a cross-platform mobile app to collect background and delivery-related information. The collected data were transferred to the cloud server for further analysis. The use of such a home-based system during pregnancy and postpartum period was evaluated in a previous study [26]. The findings of this pilot study indicated the feasibility of the study, robustness of the system, and reliability of the collected HRV parameters.

Participants and Recruitment

Women with singleton pregnancies who were at 12 to 15 weeks of gestation were recruited from southwest Finland. Women with both high-risk and low-risk pregnancies were recruited. Women with high-risk pregnancies were required to have a history of preterm birth (22-36 weeks of gestation) or late miscarriage (12-22 weeks of gestation). Women with low-risk pregnancies were required to have a history of full-term uncomplicated pregnancy and no pregnancy loss. All eligible participants had to be aged ≥18 years, understand Finnish, and have a smartphone running Android or iOS. The recruitment goal for each group (high risk and low risk) was 30 participants, for a targeted total of 60 participants.

Recruitment was performed via advertisements on social media and in maternity clinics from January 2019 to March 2021. The researcher scheduled face-to-face meetings with eligible pregnant women. During the meetings, the pregnant women were informed about the objective of the study. After the participants provided written informed consent, they were provided with a smartwatch and instructions. Moreover, our customized cross-platform mobile app was installed on their smartphone. Participants were asked to wear the smartwatch continuously during pregnancy and for 3 months after delivery. A total of 62 women were recruited (n=32, 52% in the high-risk group and n=30, 48% in the low-risk group), but 13% (4/32) of the women in the high-risk group withdrew from the study. Finally, all participants in both the high-risk and low-risk pregnancy groups were combined into 1 group for the analyses because there were no significant differences in their HRV trends. Table 1 shows the participants' background information.

Table 1. Participants' background information (n=58).

Parameters	Values
Age (years), mean (SD)	31.9 (4.9)
BMI (kg/m^2), mean (SD)	25.98 (5.96)
Marital status, n (%)	
Married or cohabitation	57 (98)
Other	1 (2)
Work status, n (%)	
Working	44 (76)
Student	7 (12)
Unemployed	1 (2)
Other	6 (10)
Education, n (%)	
High school	24 (41)
College	18 (31)
University	16 (28)
Pregnancy planned, n (%)	
Yes	53 (91)
No	5 (9)
Duration of pregnancy at recruitment (week+day), mean (SD)	14+3 (1+4)
Duration of pregnancy at birth (week+day), mean (SD)	36+4 (9+6)
Mode of delivery, n (%)	
Vaginal	48 (83)
Cesarean	10 (17)
Infant birth weight (g), mean (SD)	3532.7 (561.2)

Data Collection

Data collection was performed using the Samsung Gear Sport smartwatch and a cross-platform mobile app. The lightweight smartwatch was chosen based on its onboard sensors, battery life, configurability, internal memory, and processing unit. Moreover, the smartwatch provided access to raw PPG signals and enabled continuous data collection. The watch runs the Tizen operating system, which is open source. The open-source platform enabled us to develop customized data collection applications for the watch. We used the smartwatch to collect 12 minutes of PPG signals every 2 hours at a sampling frequency of 20 Hz. The setup was selected to enable battery life of 2 to 3 days after each full recharge [26]. The collected data were stored on the internal storage of the smartwatch. We also developed an application for the smartwatch to send the data manually through the Wi-Fi connection to our cloud server. We asked the participants to upload their data regularly. The internal storage was sufficient to store the collected data for 2 months. However, we sent notifications to the participants if they did not upload the data for 2 weeks. In addition, a cross-platform mobile app collected background information about pregnancy and infant-related data after delivery.

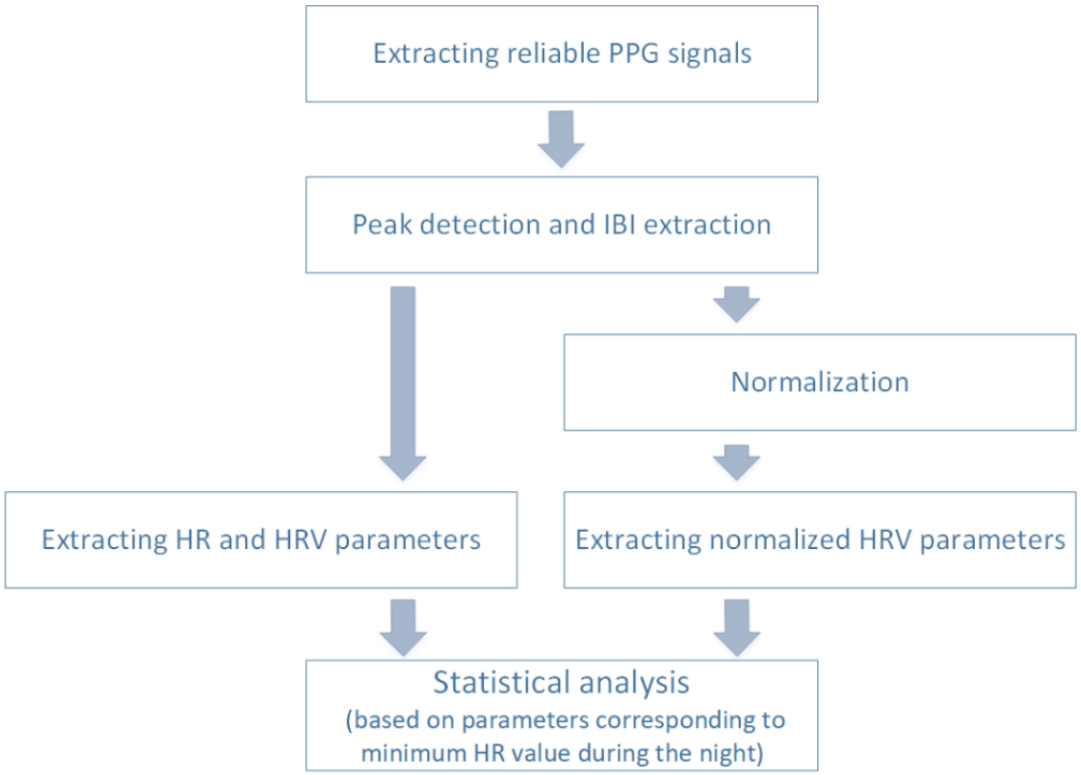
We collected PPG signals for extracting HRV parameters. Nighttime PPG data were used in this study to extract the HRV trends during pregnancy and postpartum period. Daytime PPG data were discarded as participants were involved in various activities and environments during the day, making PPG signals unreliable owing to movement artifacts and environmental noises.

Data Analysis

Overview

We analyzed the collected data on the cloud server. Data analysis included several steps, as shown in Figure 1. First, we identified and extracted reliable PPG signals. Then, a peak detection method was used to extract the peaks and IBIs. In the next step, we normalized the reliable signals to reduce the effect of HR changes on HRV parameters (refer to the *Parameter Normalization* section). We used reliable signals and normalized signals to extract reliable HR and HRV parameters. Then, we leveraged the HRV parameters during the nighttime, when resting HR has the lowest value and artifacts would be minimum to analyze HRV trends during pregnancy and the postpartum period. Finally, we used hierarchical linear mixed (HLM) models to analyze the trends of HRV parameters during pregnancy and the postpartum period.

Figure 1. Data analysis pipeline. HR: heart rate; HRV: HR variability; IBI: interbeat interval; PPG: photoplethysmogram.



Extracting Reliable Signals

PPG is a noninvasive optical method for extracting HR and HRV parameters. This method is easy to implement, and many wearables include PPG sensors. However, this method is susceptible to environmental noise and motion artifacts. Such noise can affect the quality of the signal and the accuracy of the analysis [27,28]. Therefore, unreliable signals must be detected and discarded. Reliable signals are expected to have similar waveforms, whereas unreliable signals have diverse waveforms as they are affected by different motion artifacts and environmental noises [29]. First, we extracted several morphological features from the signals and heart cycles. Then, we chose skewness, kurtosis, approximate entropy, Shannon entropy, and spectral entropy based on the scoring value for clustering [29]. We trained a support vector machine classifier using these features to distinguish between the reliable and unreliable PPG signals. Using this classifier, we discarded unreliable signals and used reliable signals in our analysis.

Peak Detection and IBI Extraction

We used a bandpass filter with cutoff frequencies of 0.7 Hz and 3.5 Hz to enhance PPG signals by filtering noises that are not in human HR ranges. Then, we used the peak detection method based on the moving average, as described in [30], to find the peaks that correspond to heartbeats. The method is enabled by an adaptive threshold, which considers the variations in the morphology and amplitude of PPG signals [31]. Then, the detected peaks were used to extract IBI, which is the interval between 2 consecutive peaks. In the error detection phase, IBIs that deviated >30% from the mean IBIs of the segments (5 minutes of signals) were removed from the IBI lists. We leveraged the HeartPy library in Python for this analysis [31].

Parameter Normalization

Studies suggest that HRV parameters are significantly associated with average HR [32]. Therefore, changes in HRV parameters result from changes in HR or HR variation [32]. In addition, several studies have found that HR increases during pregnancy [1,2]. We normalized the HRV parameters based on HR to cancel the inevitable effect of HR changes on HRV parameters. Moreover, normalization is required to compare the HRV parameters of different people, because each person's HR and resting HR are unique. Normalization was performed by dividing the IBIs by the corresponding average IBI values [33].

HR and HRV Extraction

We used the detected peaks to extract HR, the number of peaks (heartbeats) per minute. HRV parameters were obtained by extracting the variation in IBIs and normalized IBIs in the PPG signals. We used short-term HRV analysis, which requires 5-minute recordings of reliable PPG signals [25,34]. We leveraged the IBIs in each 5-minute window of reliable PPG signals to extract time-domain HRV parameters, including average of normal IBIs (AVNN), RMSSD, and SD of normal IBIs (SDNN), and frequency-domain parameters, including LF (power in low-frequency range), HF (power in high-frequency range), and LF/HF (LF to HF ratio). These parameters can be reliably extracted at a sampling frequency of 20 Hz [35].

The time-domain HRV parameters show the variation in IBIs during the monitoring period. The SDNN in the 5-minute resting measurements mainly shows the variation in parasympathetically-mediated respiratory sinus arrhythmia. RMSSD reflects the variation in successive normal IBIs. Moreover, RMSSD is the most commonly used HRV parameter for investigating vagal changes [25].

The frequency-domain HRV parameters decompose the HRV to different frequency ranges. LF can be produced by both parasympathetic and sympathetic systems. HF reflects the parasympathetic nervous system and is correlated with RMSSD [25]. Moreover, LF/HF shows the ratio of LF to HF.

We also computed the corresponding normalized parameters discussed in the *Parameter Normalization* section. The normalization resulted in normalized SDNN (nSDNN), normalized RMSSD (nRMSSD), normalized LF (nLF), normalized HF (nHF), and nLF/nHF. The HRV parameters used in this study and their definitions are presented in Table 2.

Table 2. Heart rate variability parameters.

Parameter and types	Unit	Description
Time-domain		
AVNN	ms	Average of normal IBIs[a]
SDNN	ms	SD of normal IBIs
nSDNN	ms	SD of normalized IBIs
RMSSD	ms	Square root of the mean of the sum of the squares of differences between adjacent normal IBIs
nRMSSD	ms	Square root of the mean of the sum of the squares of differences between adjacent normalized IBIs
Frequency-domain		
LF	ms^2	Power in low-frequency range (0.04-0.15 Hz)
nLF	ms^2	Power in low-frequency range (0.04-0.15 Hz) in normalized IBIs
HF	ms^2	Power in high-frequency range (0.15-0.4 Hz)
nHF	ms^2	Power in high-frequency range (0.15-0.4 Hz) in normalized IBIs
LF/HF	N/A[b]	Ratio of LF to HF
nLF/nHF	N/A	Ratio of nLF to nHF

[a]IBI: interbeat interval.

[b]N/A: not applicable.

Statistical Analysis

We used HLM models [36,37] to analyze the trends in the HRV parameter. The HLM method considers within-person and between-person changes. The HLM model supports multilevel statistical analysis when we have repeated measurements that are not independent and can correctly model correlated errors [36]. This model assumes a linear relationship between dependent and independent variables. It also enables hierarchical analysis and comparison of continuous dependent variables during different time frames (eg, before-after studies) [36].

We used the HLM models to evaluate the changes in HRV parameters during monitoring. We investigated trends in the second trimester (16-28 weeks of gestation), third trimester (29-40 weeks of gestation), and postpartum period (12 weeks after delivery). In the HLM models, HRV parameters were treated as dependent variables and time (days) was the independent variable. Therefore, the HLM model investigated HRV trends in the desired period while considering the dependency of the measurements from individual participants. We also used background parameters including age, BMI, and education level as controlling factors and analyzed their correlation with the HRV trends. We removed occupation, which is correlated with education level; planned pregnancy; and marital status, as there were few samples of unplanned pregnancy and not married or cohabitation marital status.

We included data from all the participants in the second trimester and postpartum period analyses. However, we removed the data of 12% (7/58) of the participants from the third trimester analysis owing to preterm births. It should be noted that 43% (25/58) of the participants had term delivery before the 40th gestational week and 21% (12/58N) of the participants had delivery after the 40th gestational week.

We also used HLM models to compare the second trimester with the third trimester, the second trimester with the postpartum period, and the third trimester with the postpartum period. For these analyses, we used HRV parameters as the dependent variable, time (days) as a within-person independent variable, and 1 binary independent between-person variable showing the comparing periods. We also included age, BMI, and education level in the analysis. The HLM model enabled us to perform this multilevel statistical analysis, comparing HRV trends between 2 time frames. Similarly, we included only the participants with term birth in the third trimester. All the analyses were performed using the statsmodels library in Python [38].

Ethics Approval

This study received ethics approval from the Ethics Committee of the Hospital District of Southwest Finland (Dnro: 1/1801/2018). Written informed consent was obtained from all the participants.

Results

Overview

In this section, we present the HR and HRV parameters collected during the second and third trimesters and the 3-month postpartum period. We also present the correlation between HR and HRV trends and age, BMI, and education level. Then, we compare the trends of HRV parameters between the second and third trimesters and between each trimester and the 3-month postpartum period. Data from 58 women were included in this study. The results include 166.5 (SD 46.9) reliable night data per participant during the study, with a total of 9826 night data (70% of possible data) included in this study.

Second Trimester

A total of 77.70% (4123/5306) of reliable night data were collected in the second trimester. On average, each participant had 69.9 (SD 15.1) reliable night data in the second trimester. HLM model results showed that HR increased significantly, whereas the time-domain parameters (AVNN, SDNN, nSDNN, RMSSD, and nRMSSD) and the frequency-domain parameters (LF, HF, and nHF) decreased significantly during the second trimester. In addition, the results showed no significant association of BMI and education level with HR and HRV trends in the second trimester. However, there was a significant association between age and nSDNN, nRMSSD, HF, and LF/HF. Increase in age was associated with a slight decrease in nSDNN, nRMSSD, and HF and a slight increase in LF/HF. Tables 3 and 4 show the intercept; slope of changes; association of age, BMI, and education level with trends; and the average HR and HRV parameters at the end of the second trimester.

Table 3. HR[a] and time-domain HR variability trends during the second and third trimesters and the postpartum period.

Periods and variables	HR	AVNN[b]	SDNN[c]	nSDNN[d]	RMSSD[e]	nRMSSD[f]
Second trimester						
Intercept (*P* value)	62.736 (<.001)	916.443 (<.001)	84.023 (<.001)	9.079 (<.001)	89.293 (<.001)	9.986 (<.001)
Slope (*P* value)	0.045 (<.001)	–0.585 (<.001)	–0.082 (<.001)	–0.006 (<.001)	–0.103 (<.001)	–0.007 (<.001)
Age, coefficient (*P* value)	0 (.99)	0.748 (.77)	–0.737 (.09)	–0.080 (.047)	–0.774 (.07)	–0.091 (.03)
BMI, coefficient (*P* value)	0.141 (.19)	–1.211 (.38)	–0.059 (.80)	–0.002 (.93)	–0.001 (.99)	0.002 (.93)
Education level, coefficient (*P* value)	–0.816 (.46)	14.151 (.32)	1.801 (.47)	0.112 (.64)	2.440 (.29)	0.189 (.41)
Final values, mean (SD)	71.1 (7.08)	853 (85)	53.2 (18.05)	6.2 (1.8)	58.7 (22.1)	6.7 (2.1)
Third trimester						
Intercept (*P* value)	81.324 (<.001)	708.181 (<.001)	69.787 (<.001)	9.237 (<.001)	70.275 (<.001)	8.366 (<.001)
Slope (*P* value)	–0.025 (.006)	0.345 (.002)	0.084 (<.001)	0.007 (<.001)	0.071 (<.001)	0.006 (.97)
Age, coefficient (*P* value)	–0.131 (.55)	1.905 (.45)	–1.063 (.02)	–0.135 (.002)	–0.984 (.009)	–0.092 (.20)
BMI, coefficient (*P* value)	0 (.97)	0.278 (.83)	–0.049 (.83)	–0.009 (.68)	–0.003 (.99)	–0.021 (.57)
Education level, coefficient (*P* value)	–1.540 (.19)	17.459 (.20)	5.678 (.01)	0.462 (.049)	6.755 (.001)	1.044 (.006)
Final values, mean (SD)	68.6 (7.2)	886.9 (99.5)	65.8 (24.7)	7.2 (2.8)	65.5 (26.6)	7.5 (3.1)
Postpartum period						
Intercept (*P* value)	47.237 (<.001)	1216.255 (<.001)	115.321 (<.001)	9.481 (<.001)	135.233 (<.001)	11.620 (<.001)
Slope (*P* value)	–0.009 (.37)	0.130 (.46)	0.001 (.95)	–0.001 (.69)	–0.037 (.12)	–0.004 (.01)
Age, coefficient (*P* value)	0.289 (.099)	–4.729 (.14)	–1.320 (.009)	–0.086 (.03)	–1.857 (.001)	–0.139 (.002)
BMI, coefficient (*P* value)	0.097 (.33)	–1.258 (.48)	–0.161 (.56)	–0.003 (.90)	–0.154 (.62)	–0.007 (.76)
Education level, coefficient (*P* value)	–0.300 (.77)	5.721 (.76)	–0.122 (.97)	–0.053 (.82)	1.168 (.71)	0.061 (.81)
Final values, mean (SD)	58.5 (5.9)	1037.7 (105.3)	67.3 (21.6)	6.5 (2)	69 (24.2)	6.7 (2.2)

[a]HR: heart rate.

[b]AVNN: average normal interbeat intervals.

[c]SDNN: SD of normal interbeat intervals.

[d]nSDNN: normalized SDNN.

[e]RMSSD: root mean square of the successive difference of normal interbeat intervals.

[f]nRMSSD: normalized RMSSD.

Table 4. Trends of frequency-domain heart rate variability parameters during the second and third trimesters and the postpartum period.

Periods and variables	LF[a]	nLF[b]	HF[c]	nHF[d]	LF/HF	nLF/nHF
Second trimester						
Intercept (*P* value)	1045.893 (<.001)	2.422 (<.001)	2990.343 (<.001)	4.727 (<.001)	–0.248 (<.001)	0.565 (.002)
Slope (*P* value)	–3.109 (<.001)	0.001 (.72)	–4.224 (<.001)	–0.007 (.003)	–0.061 (.10)	0.001 (.007)
Age, coefficient (*P* value)	–6.285 (.62)	–0.031 (.17)	–44.543 (.045)	–0.046 (.32)	2.775 (.001)	–0.003 (.57)
BMI, coefficient (*P* value)	2.167 (.75)	0.012 (.33)	–9.352 (.42)	0.035 (.17)	0.766 (.12)	0.002 (.48)
Education level, coefficient (*P* value)	48.903 (.50)	0.173 (.17)	71.438 (.55)	0.017 (.95)	–3.943 (.43)	0.033 (.23)
Final values, mean (SD)	677.7 (517.4)	2.3 (1.4)	1085.2 (1220.7)	3.7 (1.9)	0.83 (0.55)	0.8 (0.4)
Third trimester						
Intercept (*P* value)	1008.979 (.03)	4.248 (.001)	1743.362 (.01)	6.166 (.001)	0.67 (.84)	0.995 (.003)
Slope (*P* value)	0.424 (.44)	0.006 (.09)	2.767 (.007)	0.004 (.21)	–0.097 (.11)	0.001 (.28)
Age, coefficient (*P* value)	–19.456 (.12)	–0.070 (.03)	–40.218 (.04)	–0.098 (.054)	2.440 (.008)	–0.004 (.68)
BMI, coefficient (*P* value)	1.655 (.80)	–0.018 (.28)	–5.994 (.53)	–0.001 (.97)	0.367 (.44)	–0.006 (.17)
Education level, coefficient (*P* value)	92.269 (.17)	0.238 (.17)	221.413 (.02)	0.261 (.35)	–5.113 (.28)	0.030 (.51)
Final values, mean (SD)	926.9 (956.7)	2.6 (1.9)	1607.7 (2156.1)	4.4 (2.5)	0.77 (0.40)	0.7 (0.4)
Postpartum period						
Intercept (*P* value)	2124.653 (.002)	2.486 (.09)	4969.652 (<.001)	6.670 (.001)	–0.397 (.27)	0.416 (.02)
Slope (*P* value)	2.415 (.05)	–0.005 (.40)	–0.682 (.63)	–0.001 (.77)	0.234 (.004)	–0.002 (.05)
Age, coefficient (*P* value)	–29.427 (.14)	–0.012 (.78)	–94.705 (<.001)	–0.054 (.33)	3.498 (<.001)	0.006 (.25)
BMI, coefficient (*P* value)	0.327 (.98)	0.023 (.33)	–13.985 (.32)	0.022 (.47)	0.528 (.35)	0.002 (.40)
Education level, coefficient (*P* value)	–50.704 (.64)	0.279 (.24)	22.307 (.88)	–0.291 (.36)	–5.332 (.35)	0.053 (.06)
Final values, mean (SD)	1307.9 (920.6)	2.5 (2.5)	1486.4 (1327.6)	5.3 (2.8)	1 (0.48)	0.6 (0.3)

[a]LF: low-frequency power.

[b]nLF: normalized LF.

[c]HF: high-frequency power.

[d]nHF: normalized HF.

Third Trimester

During the third trimester, 70.25% (2716/3866) of reliable night PPG data were collected. Each participant had an average of 53.2 (SD 15.1) reliable night PPG data in the third trimester. The HLM models show that HR decreased significantly, whereas the time-domain parameters (AVNN, SDNN, nSDNN, and RMSSD) and frequency-domain parameter (HF) increased significantly during the third trimester (refer to Tables 3 and 4 for details). Moreover, the results indicated that high education level was associated with high SDNN, nSDNN, RMSSD, nRMSSD, and HF. It also showed that increase in age was associated with a slight decrease in SDNN, nSDNN, RMSSD, nLF, and HF and a slight increase in LF/HF.

Considering both trimesters as a whole, the models indicated that HR significantly increased (*P*<.001), whereas AVNN, SDNN, RMSSD, LF, and HF decreased during pregnancy (with *P*<.001, *P*=.04, *P*=.001, *P*=.44, *P*=.44, respectively). However, during the last weeks of pregnancy, starting from pregnancy week 35, HR began to decrease and HRV parameters (AVNN, SDNN, RMSSD, LF, and HF) began to increase, but they did not reach the level of pregnancy week 16 before the delivery.

Postpartum Period

In the postpartum period, 62.05% (2987/4814) of reliable night PPG data were collected from the participants. Each participant had an average of 53.4 (SD 19.7) reliable night data in this period. During the first 12 weeks after delivery, the time-domain parameter (nRMSSD) decreased slightly and the frequency-domain parameter (LF/HF) increased slightly (Tables 3 and 4). The results indicated that increase in age was associated with a slight decrease in SDNN, nSDNN, RMSSD, nRMSSD, and HF and a slight increase in LF/HF. Moreover, the results showed no significant correlation between the duration of pregnancy and trends of HR and HRV parameters during postpartum period. Figures 2 and 3 represent the trends of HR, time-domain, and frequency-domain HRV parameters during pregnancy and the 3-month postpartum period.

Figure 2. Weekly mean and 95% CI of heart rate (HR) and time-domain HR variability parameters during pregnancy and postpartum period. The number of participants with reliable data per week is also indicated. The vertical line indicates pregnancy week 28 and separates the second and third trimesters. AVNN: average normal interbeat intervals; RMSSD: root mean square of the successive differences of normal interbeat intervals; nRMSSD: normalized RMSSD; SDNN: SD of normal interbeat intervals. nSDNN: normalized SDNN.

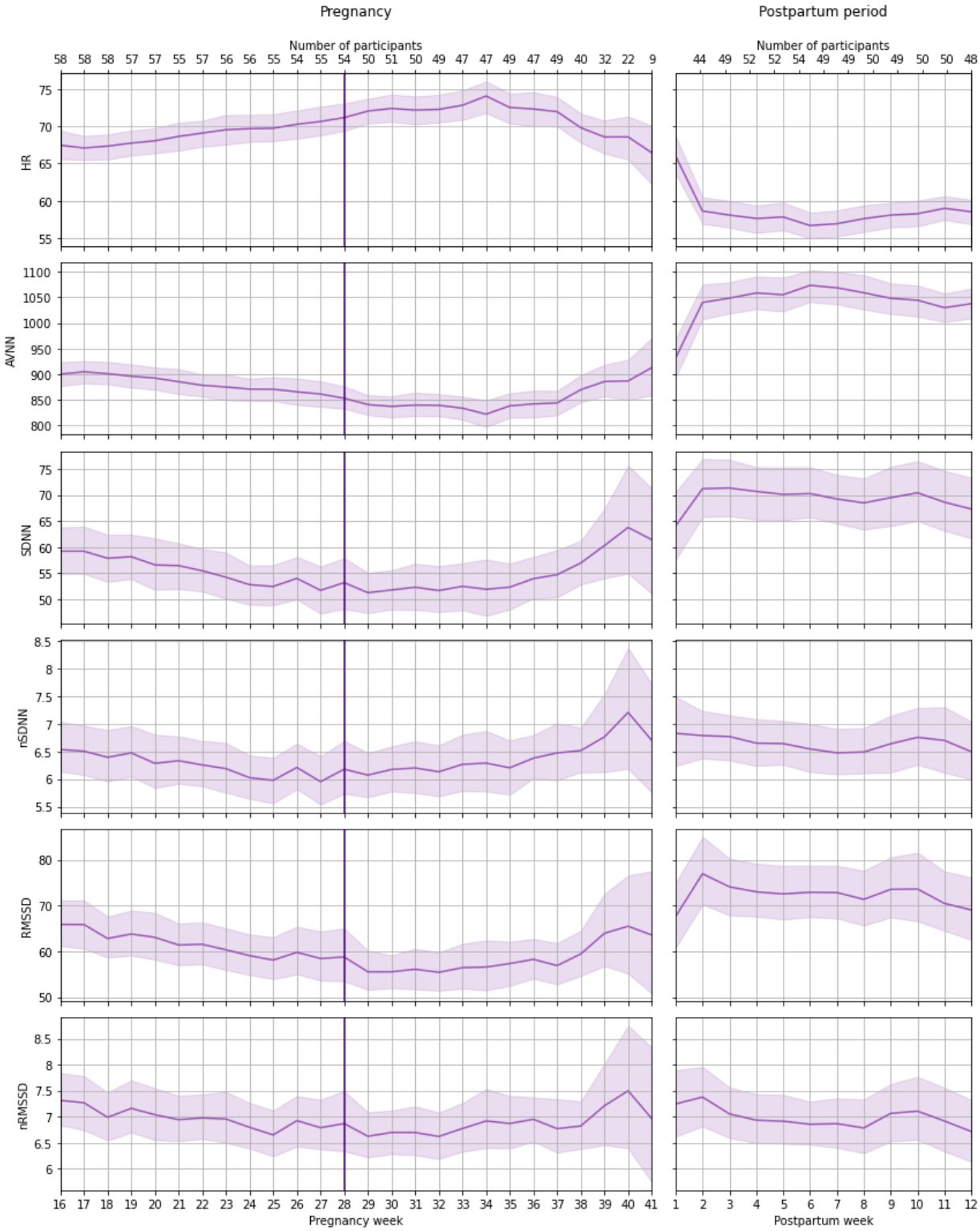
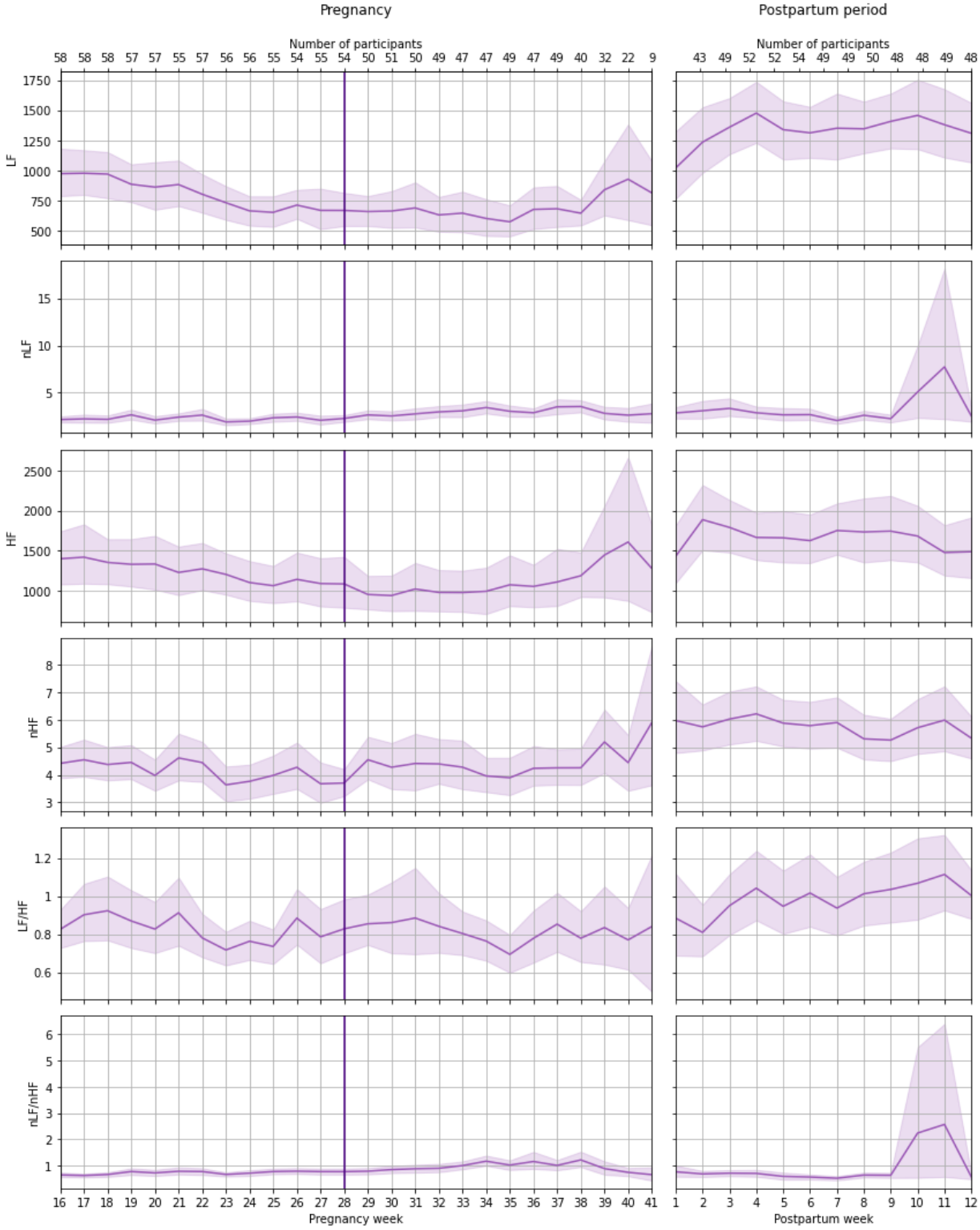


Figure 3. Weekly mean and 95% CI of frequency-domain heart rate variability parameter during pregnancy and postpartum period. The number of participants with reliable data per week is also indicated. The vertical line indicates pregnancy week 28 and separates the second and third trimesters. HF: high-frequency power; LF: low-frequency power; nHF: normalized HF; nLF: normalized LF.



Comparison of Trends in HRV Parameters Among the Second Trimester, Third Trimester, and 3-Month Postpartum Period

We compared the trends of HRV parameters among the second trimester, third trimester, and postpartum period.

The Second and Third Trimesters

The time-domain HRV parameters, including AVNN, SDNN, nSDNN, and RMSSD, and some frequency-domain parameters, including LF, HF, and nHF, were slightly higher in the third trimester than in the second trimester. Moreover, HR was slightly high in the second trimester.

The average increase in HR from the beginning of the second trimester to week 34, when HR reached its highest level was 6.58 beats per minute (bpm). Time-domain parameters, AVNN, SDNN, nSDNN, RMSSD, and nRMSSD, were 77.3, 7.8, 0.4, 9.5, 0.5 milliseconds lower, respectively, at the beginning of the third trimester than the beginning of the second trimester. In addition, the slope of per-day changes in HR was slightly lower, whereas this slope was slightly higher for AVNN, SDNN, nSDNN, RMSSD, nRMSSD, LF, HF, and nHF in the third trimester than in the second trimester (Tables 5 and 6). Moreover, the decrease in HRV trends, including SDNN, nSDNN, RMSSD, nRMSSD, and HF, in the third trimester compared with the second trimester was slightly higher in younger women than in older women. However, in this comparison, the difference in LF/HF slightly increased with increase in age.

Table 5. Comparison of HR[a] and time-domain HR variability parameters between the second and third trimesters, second trimester and postpartum period, and third trimester and postpartum period.

Comparisons and variables.	HR	AVNN[b]	SDNN[c]	nSDNN[d]	RMSSD[e]	nRMSSD[f]
Second trimester and third trimester						
Slope (*P* value)	−0.080 (<.001)	1.052 (<.001)	0.149 (<.001)	0.010 (<.001)	0.168 (<.001)	0.011 (<.001)
Intercept (*P* value)	6.128 (<.001)	−77.279 (<.001)	−7.793 (<.001)	−0.377 (<.001)	−9.535 (<.001)	−0.529 (<.001)
Age (years), coefficient (*P* value)	0.015 (.93)	0.017 (.99)	−0.859 (.03)	−0.090 (.02)	−1.097 (.006)	−0.119 (.003)
BMI (kg/m^2), coefficient (*P* value)	0.125 (.23)	−1.316 (.32)	−0.101 (.65)	−0.007 (.75)	−0.071 (.74)	−0.002 (.94)
Education level, coefficient (*P* value)	−1.396 (.20)	19.196 (.16)	2.257 (.33)	0.153 (.50)	3.689 (.10)	0.274 (.23)
Second trimester and postpartum period						
Slope (*P* value)	−0.061 (<.001)	0.800 (<.001)	0.079 (<.001)	0.004 (.009)	0.077 (<.001)	0.003 (.02)
Intercept (*P* value)	−8.135 (<.001)	133.327 (<.001)	10.458 (<.001)	0.235 (.002)	7.776 (<.001)	−0.062 (.32)
Age (years), coefficient (*P* value)	0.071 (.68)	−1.067 (.68)	−0.843 (.046)	−0.086 (.02)	−1.208 (.009)	−0.114 (.005)
BMI (kg/m^2), coefficient (*P* value)	0.122 (.22)	−1.202 (.41)	−0.047 (.84)	0.005 (.83)	−0.046 (.86)	0.002 (.94)
Education level, coefficient (*P* value)	−0.780 (.44)	11.211 (.45)	1.493 (.55)	−0.005 (.98)	2.057 (.42)	0.138 (.55)
Third trimester and postpartum period						
Slope (*P* value)	0.026 (<.001)	−0.306 (.003)	−0.101 (<.001)	−0.008 (<.001)	−0.102 (<.001)	−0.009 (<.001)
Intercept (*P* value)	−14.686 (<.001)	212.495 (<.001)	19.492 (<.001)	0.691 (<.001)	18.524 (<.001)	0.541 (<.001)
Age (years), coefficient (*P* value)	0.121 (.47)	−1.801 (.47)	−0.921 (.02)	−0.080 (.03)	−1.142 (.002)	−0.108 (.003)
BMI (kg/m^2), coefficient (*P* value)	0.097 (.29)	−1.538 (.27)	−0.132 (.55)	0.001 (.98)	−0.193 (.36)	−0.010 (.64)
Education level, coefficient (*P* value)	−1.330 (.16)	15.663 (.28)	2.892 (.21)	0.243 (.25)	4.053 (.06)	0.381 (.08)

[a]HR: heart rate.

[b]AVNN: average normal interbeat intervals.

[c]SDNN: SD of normal interbeat intervals.

[d]nSDNN: normalized SDNN.

[e]RMSSD: root mean square of the successive difference of normal interbeat intervals.

[f]nRMSSD: normalized RMSSD.

Table 6. Comparison of frequency-domain heart rate variability parameters between the second and third trimesters, second trimester and postpartum period, and third trimester and postpartum period.

Comparisons and variables	LF[a]	nLF[b]	HF[c]	nHF[d]	LF/HF	nLF/nHF
Second trimester and third trimester						
Slope (*P* value)	2.776 (<.001)	0.003 (.29)	5.628 (<.001)	0.010 (.002)	−0.061 (.25)	0 (.70)
Intercept (*P* value)	−248.629 (<.001)	0.396 (<.001)	−371.246 (<.001)	−0.419 (.005)	0.338 (.89)	0.172 (<.001)
Age (years), coefficient (*P* value)	−12.053 (.29)	−0.046 (.06)	−58.317 (.004)	−0.063 (.13)	0.027 (.003)	−0.003 (.61)
BMI (kg/m^2), coefficient (*P* value)	−0.220 (.97)	0.003 (.82)	−10.722 (.34)	0.023 (.34)	0.005 (.30)	0 (.93)
Education level, coefficient (*P* value)	42.367 (.52)	0.171 (.22)	105.380 (.36)	0.051 (.84)	−0.041 (.43)	0.023 (.46)
Second trimester and postpartum period						
Slope (*P* value)	5.261 (<.001)	−0.007 (.045)	3.748 (<.001)	0.002 (.57)	0.287 (<.001)	−0.003 (<.001)
Intercept (*P* value)	194.724 (<.001)	0.754 (<.001)	311.128 (<.001)	1.073 (<.001)	−1.367 (.63)	0.061 (.16)
Age (years), coefficient (*P* value)	−12.415 (.36)	−0.025 (.36)	−64.949 (.004)	−0.048 (.28)	3.152 (.001)	0.001 (.85)
BMI (kg/m^2), coefficient (*P* value)	1.922 (.80)	0.028 (.08)	−8.505 (.49)	0.030 (.23)	0.706 (.16)	0.003 (.27)
Education level, coefficient (*P* value)	36.128 (.65)	0.223 (.16)	69.474 (.58)	−0.013 (.96)	−3.490 (.48)	0.066 (.01)
Third trimester and postpartum period						
Slope (*P* value)	1.773 (.05)	−0.015 (.002)	−3.219 (.001)	−0.007 (.12)	0.377 (<.001)	−0.005 (<.001)
Intercept (*P* value)	496.288 (<.001)	0.240 (.28)	765.852 (<.001)	1.488 (<.001)	−2.648 (.40)	−0.160 (.008)
Age (years), coefficient (*P* value)	−17.497 (.22)	−0.039 (.23)	−65.002 (.001)	−0.079 (.09)	2.629 (.004)	0.002 (.79)
BMI (kg/m^2), coefficient (*P* value)	2.643 (.73)	0.011 (.56)	−13.823 (.20)	0.012 (.65)	0.530 (.32)	−0.002 (.63)
Education level, coefficient (*P* value)	18.376 (.82)	0.224 (.23)	142.030 (.19)	0.028 (.92)	−5.352 (.30)	0.045 (.28)

[a]LF: low-frequency power.

[b]nLF: normalized LF.

[c]HF: high-frequency power.

[d]nHF: normalized HF.

The Second Trimester and Postpartum Period

In the postpartum period, HR was significantly lower (on average 8.1 bpm), and the time-domain parameters, AVNN (133.3 milliseconds), SDNN (10.5 milliseconds), nSDNN (0.2 milliseconds), RMSSD (7.8 milliseconds), and frequency-domain parameters LF (195.1 square milliseconds), nLF (0.7 square milliseconds), and HF (312.4 square milliseconds) were significantly higher than those in the second trimester. The slope of changes in HR was slightly lower, whereas the slope of changes in AVNN, SDNN, nSDNN, RMSSD, nRMSSD, LF, nLF, HF, LF/HF, and nLF/nHF was higher than those in the second trimester (Tables 5 and 6). The difference between the trends of SDNN, nSDNN, RMSSD, nRMSSD, and HF decreased, whereas the difference in LF/HF slightly increased with increase in age.

The Third Trimester and Postpartum Period

In the postpartum period, HR was significantly lower (on average 14.7 bpm) than that at the beginning of the third trimester. However, the time-domain parameters, AVNN, SDNN, nSDNN, RMSSD, and nRMSSD, and frequency-domain parameters, LF, HF, and nHF were, on average, 212.8 milliseconds, 19.4 milliseconds, 0.7 milliseconds, 18.5 milliseconds, 0.5 milliseconds, 495 square milliseconds, 764.6 square milliseconds, and 1.5 square milliseconds higher, respectively, at the beginning of the postpartum period than at the beginning of the third trimester. The slope of changes in

AVNN, SDNN, nSDNN, RMSSD, nRMSSD, nLF, HF, and nHF was slightly higher, whereas the slope of changes in LF was slightly lower in the third trimester than in the postpartum period (Tables 5 and 6). In addition, the difference between the trends of SDNN, nSDNN, RMSSD, nRMSSD, and HF decreased, whereas the difference in LF/HF slightly increased with increase in age.

Discussion

Principal Findings

Our results show that HR increased significantly during the second trimester, whereas it slightly decreased during the third trimester. During the postpartum period, HR continued to decrease, but the reduction was not statistically significant; however, compared with pregnancy, HR was significantly low during the postpartum period. On average, HR increased by 6.6 bpm from 16 weeks to 34 weeks of gestation, after which, it started to decrease.

The trends detected in this study are consistent with the previous review by Loerup et al [2], in which the mean increase was 7.6 (95% CI 1.8-13.4) bpm from 10 weeks to 40 weeks of gestation. The increase in HR during pregnancy is considered physiological and explained by elevated blood volume, which results in increased cardiac output [7].

Regarding HRV, the time-domain parameters and their normalized values decreased significantly during the second trimester and, then, increased significantly during the third trimester. However, these parameters did not reach the level of those during the second trimester. In the postpartum period, the time-domain parameters were stable, and only nRMSSD decreased. Regarding the frequency-domain parameters, LF, HF, and nHF decreased significantly during the second trimester, whereas nLF/nHF increased slightly. During the third trimester and postpartum period, the parameters were stable, except HF, which increased, and LF/HF, which decreased slightly. Changes in HRV parameters occur owing to the pregnancy and the physiological changes in the woman's body [1]. The trend in HRV parameters during pregnancy was decreasing, with values returning to normal after delivery [3-5].

The results indicated that BMI is not significantly associated with HRV trends. In addition, younger women had higher nSDNN, nRMSSD, and HF in the second trimester and lower SDNN, nSDNN, RMSSD, nRMSSD, and HF and slightly lower LF/HF during the third trimester and postpartum period than older women. These results are consistent with previous studies showing a negative correlation between age and HRV parameters [25]. Furthermore, the results showed that more educated women had higher SDNN, nSDNN, RMSSD, nRMSSD, and HF during the third trimester than the less educated women, which may indicate low stress level in highly educated women. Previous studies also showed that highly educated people experienced low stress in a stressful situation [39], and low education level is identified as a determinant of stress during pregnancy [40].

Comparison With Previous Work

To the best of our knowledge, this is the first paper that describes and evaluates HR and HRV parameters measured using PPG signals continuously during pregnancy and the postpartum period in participants' normal daily lives during the night. Previous studies have been limited to a few samples assessed in controlled environments during pregnancy and the postpartum period. Only Stein et al [1] performed a study in free-living conditions; they measured HRV using Holter ECG for 24-hour periods during pregnancy.

Continuously measured HR followed the physiological trends, increasing as the pregnancy proceeded and returning to normal during the postpartum period. However, it is notable that in this study, we did not measure HR levels before pregnancy; thus, it is not possible to confirm whether HR levels returned to prepregnancy levels during the 3-month follow-up. Several studies have shown that HR increases during pregnancy [7,15,21,24] and decreases again during the postpartum period [4]. In this study, the detected increase during pregnancy followed the results of the meta-analysis, which included >10,000 HR measurements from >8000 women [2]. The small difference may be explained by the small sample size in this study and differences in the measurement periods, as our study measurements started at gestational week 16 and the meta-analysis started from gestational week 10 [2]. However, it is suggested that most of the changes in cardiac autonomic modulation occur within the first weeks after conception [1]; thus, in this study, we were not able to detect the early changes.

Interestingly, according to our study, HR was the highest during pregnancy week 34 and started to decrease thereafter. In many previous studies [1,23,24], the last measurement points were before week 36; thus, the decrease in HR at the end of the pregnancy may not have been captured. On the basis of the meta-analysis by Loerup et al [2], a few previous studies have shown a slight reduction in HR at the end of pregnancy. However, most studies show a continuous increase during pregnancy.

The results showed that all the time-domain HRV parameters measured in this study and the frequency-domain parameters (LF and HF) decreased during the second trimester. Furthermore, most of the measured parameters (AVNN, SDNN, RMSSD, LF, and HF) showed decreasing trends throughout pregnancy. On the basis of previous studies with intermittent measurements, the trends of different HRV parameters decreased during the course of pregnancy [1,7,9,11,15,22-24]. Some studies also found increasing trends, for example, in LF [23], and some did not find any significant changes during pregnancy [16,21]. These conflicting results are partly owing to different methodological choices, such as limited HRV recordings and a small number of participants, but they also reflect the challenges of measuring and interpreting HRV [25].

On the basis of the continuous measurements in our study, a change from a decreasing trend to a slightly increasing trend in HRV parameters was observed during the last weeks of pregnancy, starting at week 35. Most previous studies included very few or no measurements of HRV after pregnancy week 35; therefore, the change has probably not been detected

[1,9,22-24]. Long intervals (eg, weeks) between successive HRV measurements restrict the findings regarding fine-grained trends at the end of pregnancy. Continuous measurements provide opportunities to detect small changes also. In Finland, pregnant women are entitled to maternity leave starting 5 to 8 weeks before the estimated delivery date [41]. Thus, we could speculate that one explanation for the changes in HR and HRV parameters around gestational weeks 34 to 35 could be the beginning of the maternity leave. Maternity leave allows, for example, a woman to modify her daily rhythm, and therefore, the level of stress may decrease. However, this issue requires more research in the future. It is also notable that the data for this study were collected partly during the COVID-19 pandemic and the first wave of restrictions, which may have affected the behavior and, by implication, changes in the physiological parameters of the participating women [42].

Some HRV parameters were negatively associated with age, whereas LF/HF was positively associated with age, in our study. Several studies have shown similar correlations between age and HRV parameters [25,43]. Changes in HRV parameters are also associated with stress, as HRV represents the ability of the heart to respond to a variety of psychological and environmental stimuli [44]. Although cardiovascular changes during pregnancy are physiological, Klinkenberg et al [13] suggested that psychosocial stress also affects HRV parameters in pregnant women. Low values of SDNN, RMSSD, and HF and high values of LF and LF/HF may indicate mental stress [14]. Our results showed a positive correlation between HRV parameters and education level in the third trimester, which may indicate low stress level in highly educated women [39,40]. However, interpreting HRV parameters regarding stress is difficult owing to physiological changes during pregnancy and the variety of potential stressors and individual stress responses [25,44].

During the postpartum period, some of the HRV parameters (SDNN, RMSSD, LF, and HF) increased as expected, as the body recovers from the pregnancy and delivery and returns to the *normal* nonpregnant state [1,5,19]. It is suggested that autonomic nervous system recovers approximately 4 months after delivery [3]. On the basis of only one HRV measurement during the third trimester and another at 3-month postpartum period, Heiskanen et al [4] found similar results regarding the frequency-domain HRV parameters; the parameters significantly increased from the third trimester to the postpartum period. They suggest that the optimal time for measuring the recovery of HRV is 6 months after childbirth; however, possible new pregnancy or the use of oral contraceptives may affect the results at that point [4]. In a recently published study by Brown et al [5], the only significant change in HRV was observed between the third trimester and 4 to 6 weeks of the postpartum period.

In this study, we continuously collected HRV data from pregnant women using an IoT-based maternal monitoring system [26]. The system could collect a considerable amount of HRV data. We were able to extract reliable data from >70% of possible nights during pregnancy and >60% of nights after delivery. The results indicate that continuous HRV monitoring with PPG signals can be used in free-living conditions during pregnancy and the postpartum period. In contrast to previous studies, our results contained fine-grained HRV data, which enabled us to investigate the trends with more granularity.

HRV monitoring during pregnancy could be used for the early detection of complications, such as gestational hypertension [6] and pre-eclampsia [8,9], as reflected in previous studies. For example, Hossen et al [8] developed a model based on frequency-domain HRV parameters to distinguish between pre-eclampsia and normal pregnancy. In addition to interesting HRV trends during pregnancy and the postpartum period, the results of this study showed the feasibility of the IoT-based system for remote HRV monitoring of maternal health. This system can be further developed to build a personalized model that uses individual parameters and normal HRV trends for early anomaly detection. This model may even provide early warning for mothers in a noninvasive and cost-efficient manner. This technology could provide a solution to support maternal health services in low- and middle-income countries. Although many other efforts are also needed [45], technology could enhance health equality between urban and rural areas.

Limitations

Women with both high-risk and low-risk pregnancies were included in the sample; however, no differences were detected in HRV parameters between the 2 groups, and therefore, the sample was considered as one group. Only nighttime data were used for the analyses; the minimum resting HR between midnight and 6 AM was used to analyze the trend of HR and corresponding HRV parameters to minimize the effect of noises and artifacts [25]. HRV was measured using PPG signals, which were collected with a frequency of 20 Hz. Therefore, our results need to be interpreted with caution, as not all HRV parameters can be obtained reliably at this frequency [26,35]. Moreover, some studies suggest that HRV changes occur mostly during early pregnancy [1,19], and these changes could not be detected in this study because data collection started at pregnancy week 16. Our future work will consider using high-frequency PPG signals to study other HRV trends. Furthermore, when the participants are involved in different activities, daytime HRV parameters would also provide interesting data if the noises and artifacts caused by movement could be removed from the data. In addition, it would be important to control the HRV analysis for various confounding factors such as medical conditions (eg, hypertension) and mental distress.

Conclusions

In this study, we conducted continuous long-term measurements of HR and HRV from pregnancy week 16 to 3 months of the postpartum period during participants' daily lives. The measurements were performed through the collection of PPG signals from wearable smartwatches. The results showed that HR and HRV mainly followed the expected and previously reported trends; HR increased and HRV parameters decreased as pregnancy proceeded, and the values returned to normal after delivery. These trends reflect the normal physiological changes during pregnancy and postpartum period. However, from the continuous measurements, we detected that HR started to decrease and HRV parameters started to increase during the last weeks of pregnancy. This issue needs more research in the future. The results also showed a positive association between

HRV parameters and education level in the third trimester. Furthermore, our results showed that using PPG signals, it is possible to follow HRV continuously in free-living conditions. Our system could be further developed and used in the future; for example, to detect abnormalities during pregnancy.

Acknowledgments

This study was partially supported by the Academy of Finland through the Supporting Lifestyle Change in Obese Pregnant Mothers through Wearable Internet-of-Things project under grants 316810 and 316811 and the US National Science Foundation under grants Wireless Innovation between Finland and United States (CNS-1702950) and Smart and Connected Communities (CNS-1831918).

Conflicts of Interest

None declared.

References

1. Stein PK, Hagley MT, Cole PL, Domitrovich PP, Kleiger RE, Rottman JN. Changes in 24-hour heart rate variability during normal pregnancy. Am J Obstet Gynecol 1999 Apr;180(4):978-985. [doi: 10.1016/s0002-9378(99)70670-8] [Medline: 10203667]
2. Loerup L, Pullon RM, Birks J, Fleming S, Mackillop LH, Gerry S, et al. Trends of blood pressure and heart rate in normal pregnancies: a systematic review and meta-analysis. BMC Med 2019 Sep 11;17(1):167 [FREE Full text] [doi: 10.1186/s12916-019-1399-1] [Medline: 31506067]
3. Ekholm EM, Piha SJ, Erkkola RU, Antila KJ. Autonomic cardiovascular reflexes in pregnancy. A longitudinal study. Clin Auton Res 1994 Aug;4(4):161-165. [doi: 10.1007/BF01826181] [Medline: 7849495]
4. Heiskanen N, Saarelainen H, Valtonen P, Lyyra-Laitinen T, Laitinen T, Vanninen E, et al. Blood pressure and heart rate variability analysis of orthostatic challenge in normal human pregnancies. Clin Physiol Funct Imaging 2008 Nov;28(6):384-390. [doi: 10.1111/j.1475-097X.2008.00818.x] [Medline: 18691249]
5. Brown RL, Fagundes CP, Thayer JF, Christian LM. Longitudinal changes in HRV across pregnancy and postpartum: effect of negative partner relationship qualities. Psychoneuroendocrinology 2021 Jul;129:105216. [doi: 10.1016/j.psyneuen.2021.105216] [Medline: 33964738]
6. Moors S, Staaks KJ, Westerhuis ME, Dekker LR, Verdurmen KM, Oei SG, et al. Heart rate variability in hypertensive pregnancy disorders: a systematic review. Pregnancy Hypertens 2020 Apr;20:56-68. [doi: 10.1016/j.preghy.2020.03.003] [Medline: 32179490]
7. Walther T, Wessel N, Baumert M, Stepan H, Voss A, Faber R. Longitudinal analysis of heart rate variability in chronic hypertensive pregnancy. Hypertens Res 2005 Feb;28(2):113-118. [doi: 10.1291/hypres.28.113] [Medline: 16025737]
8. Hossen A, Barhoum A, Jaju D, Gowri V, Al-Hashmi K, Hassan M, et al. Investigation of the high frequency band of heart rate variability: identification of preeclamptic pregnancy from normal pregnancy in Oman. Asian Biomed 2013;7(3):339-346. [doi: 10.3233/thc-160681]
9. Yang CC, Chao TC, Kuo TB, Yin CS, Chen HI. Preeclamptic pregnancy is associated with increased sympathetic and decreased parasympathetic control of HR. Am J Physiol Heart Circ Physiol 2000 Apr;278(4):H1269-H1273 [FREE Full text] [doi: 10.1152/ajpheart.2000.278.4.H1269] [Medline: 10749724]
10. Shea AK, Kamath MV, Fleming A, Streiner DL, Redmond K, Steiner M. The effect of depression on heart rate variability during pregnancy. A naturalistic study. Clin Auton Res 2008 Aug;18(4):203-212. [doi: 10.1007/s10286-008-0480-1] [Medline: 18592128]
11. Mizuno T, Tamakoshi K, Tanabe K. Anxiety during pregnancy and autonomic nervous system activity: a longitudinal observational and cross-sectional study. J Psychosom Res 2017 Aug;99:105-111. [doi: 10.1016/j.jpsychores.2017.06.006] [Medline: 28712414]
12. Braeken MA, Jones A, Otte RA, Widjaja D, Van Huffel S, Monsieur GJ, et al. Anxious women do not show the expected decrease in cardiovascular stress responsiveness as pregnancy advances. Biol Psychol 2015 Oct;111:83-89. [doi: 10.1016/j.biopsycho.2015.08.007] [Medline: 26316361]
13. Klinkenberg AV, Nater UM, Nierop A, Bratsikas A, Zimmermann R, Ehlert U. Heart rate variability changes in pregnant and non-pregnant women during standardized psychosocial stress. Acta Obstet Gynecol Scand 2009;88(1):77-82 [FREE Full text] [doi: 10.1080/00016340802566762] [Medline: 19023678]
14. Teckenberg-Jansson P, Turunen S, Pölkki T, Lauri-Haikala MJ, Lipsanen J, Henelius A, et al. Effects of live music therapy on heart rate variability and self-reported stress and anxiety among hospitalized pregnant women: a randomized controlled trial. Nordic J Music Ther 2019 Jan 24;28(1):7-26. [doi: 10.1080/08098131.2018.1546223]
15. Braeken MA, Jones A, Otte RA, Nyklíček I, Van den Bergh BR. Potential benefits of mindfulness during pregnancy on maternal autonomic nervous system function and infant development. Psychophysiology 2017 Feb;54(2):279-288. [doi: 10.1111/psyp.12782] [Medline: 28118687]

16. Voss A, Malberg H, Schumann A, Wessel N, Walther T, Stepan H, et al. Baroreflex sensitivity, heart rate, and blood pressure variability in normal pregnancy. Am J Hypertens 2000 Nov 01;13(11):1218-1225. [doi: 10.1016/s0895-7061(00)01199-7]
17. Solanki JD, Desai FH, Desai KH. Heart rate variability is reduced in normal pregnancy irrespective of trimester: a cross-sectional study from Gujarat, India. J Family Med Prim Care 2020 Feb;9(2):626-631 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_1123_19] [Medline: 32318393]
18. Chamchad D, Horrow JC, Nakhamchik L, Arkoosh VA. Heart rate variability changes during pregnancy: an observational study. Int J Obstet Anesth 2007 Apr;16(2):106-109. [doi: 10.1016/j.ijoa.2006.08.008] [Medline: 17270423]
19. Speranza G, Verlato G, Albiero A. Autonomic changes during pregnancy: assessment by spectral heart rate variability analysis. J Electrocardiol 1998 Apr;31(2):101-109. [doi: 10.1016/s0022-0736(98)90040-1] [Medline: 9588655]
20. Reyes-Lagos JJ, Echeverría-Arjonilla JC, Peña-Castillo M, García-González MT, Ortiz-Pedroza MD, Pacheco-López G, et al. A comparison of heart rate variability in women at the third trimester of pregnancy and during low-risk labour. Physiol Behav 2015 Oct 01;149:255-261. [doi: 10.1016/j.physbeh.2015.05.041] [Medline: 26048301]
21. Moertl MG, Ulrich D, Pickel KI, Klaritsch P, Schaffer M, Flotzinger D, et al. Changes in haemodynamic and autonomous nervous system parameters measured non-invasively throughout normal pregnancy. Eur J Obstet Gynecol Reprod Biol 2009 May;144 Suppl 1:S179-S183. [doi: 10.1016/j.ejogrb.2009.02.037] [Medline: 19285779]
22. Baumert M, Javorka M, Seeck A, Faber R, Sanders P, Voss A. Multiscale entropy and detrended fluctuation analysis of QT interval and heart rate variability during normal pregnancy. Comput Biol Med 2012 Mar;42(3):347-352. [doi: 10.1016/j.compbiomed.2011.03.019] [Medline: 21530956]
23. Garg P, Yadav K, Jaryal AK, Kachhawa G, Kriplani A, Deepak KK. Sequential analysis of heart rate variability, blood pressure variability and baroreflex sensitivity in healthy pregnancy. Clin Auton Res 2020 Oct;30(5):433-439. [doi: 10.1007/s10286-020-00667-4] [Medline: 31981003]
24. Gandhi PH, Mehta HB, Gokhale AV, Desai CB, Gokhale PA, Shah CJ. A study on cardiac autonomic modulation during pregnancy by non-invasive heart rate variability measurement. Int J Med Public Health 2014 Oct;4(4):441-445. [doi: 10.4103/2230-8598.144131]
25. Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. Front Public Health 2017 Sep 28;5:258 [FREE Full text] [doi: 10.3389/fpubh.2017.00258] [Medline: 29034226]
26. Sarhaddi F, Azimi I, Labbaf S, Niela-Vilén H, Dutt N, Axelin A, et al. Long-term IoT-based maternal monitoring: system design and evaluation. Sensors (Basel) 2021 Mar 24;21(7):2281 [FREE Full text] [doi: 10.3390/s21072281] [Medline: 33805217]
27. Elgendi M. Optimal Signal Quality Index for photoplethysmogram signals. Bioengineering (Basel) 2016 Sep 22;3(4):21 [FREE Full text] [doi: 10.3390/bioengineering3040021] [Medline: 28952584]
28. Naeini EK, Azimi I, Rahmani AM, Liljeberg P, Dutt N. A real-time PPG quality assessment approach for healthcare Internet-of-Things. Procedia Comput Sci 2019;151:551-558. [doi: 10.1016/j.procs.2019.04.074]
29. Mahmoudzadeh A, Azimi I, Rahmani AM, Liljeberg P. Lightweight photoplethysmography quality assessment for real-time IoT-based health monitoring using unsupervised anomaly detection. Procedia Comput Sci 2021;184:140-147. [doi: 10.1016/j.procs.2021.03.025]
30. van Gent P, Farah P, Nes N, van Arem B. Heart rate analysis for human factors: development and validation of an open source toolkit for noisy naturalistic heart rate data. In: Proceedings of the 6th Humanist Conference. 2018 Presented at: HUMMANIST '18; June 13-14, 2018; The Hague, The Netherlands.
31. van Gent P, Farah H, van Nes N, van Arem B. Analysing noisy driver physiology real-time using off-the-shelf sensors: heart rate analysis software from the taking the fast lane project. J Open Res Softw 2019 Oct 29;7(1):32. [doi: 10.5334/jors.241]
32. Sacha J, Barabach S, Statkiewicz-Barabach G, Sacha K, Müller A, Piskorski J, et al. How to strengthen or weaken the HRV dependence on heart rate--description of the method and its perspectives. Int J Cardiol 2013 Sep 30;168(2):1660-1663. [doi: 10.1016/j.ijcard.2013.03.038] [Medline: 23578892]
33. Gąsior JS, Sacha J, Jeleń PJ, Pawłowski M, Werner B, Dąbrowski MJ. Interaction between heart rate variability and heart rate in pediatric population. Front Physiol 2015 Dec 18;6:385 [FREE Full text] [doi: 10.3389/fphys.2015.00385] [Medline: 26733878]
34. Malik M, Bigger JT, Camm AJ, Kleiger RE, Malliani A, Moss AJ, et al. Heart rate variability: standards of measurement, physiological interpretation, and clinical use. Eur Heart J 1996 Mar 01;17(3):354-381. [doi: 10.1093/oxfordjournals.eurheartj.a014868]
35. Choi A, Shin H. Photoplethysmography sampling frequency: pilot assessment of how low can we go to analyze pulse rate variability with reliability? Physiol Meas 2017 Mar;38(3):586-600. [doi: 10.1088/1361-6579/aa5efa] [Medline: 28169836]
36. Raudenbush SW, Bryk AS. Hierarchical Linear Models: Applications and Data Analysis Methods. 2nd edition. Thousand Oaks, CA, USA: Sage Publications; 2002.
37. Bolger N, Laurenceau JP. Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research. New York, NY, USA: The Guilford Press; 2013.

38. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with Python. In: Proceedings of the 9th Python in Science Conference. 2010 Presented at: SciPy '10; June 28-30, 2010; Austin, TX, USA. [doi: 10.25080/majora-92bf1922-011]
39. Fiocco AJ, Joober R, Lupien SJ. Education modulates cortisol reactivity to the Trier Social Stress Test in middle-aged adults. Psychoneuroendocrinology 2007;32(8-10):1158-1163. [doi: 10.1016/j.psyneuen.2007.08.008] [Medline: 17920776]
40. Cardwell MS. Stress: pregnancy considerations. Obstet Gynecol Surv 2013 Feb;68(2):119-129. [doi: 10.1097/OGX.0b013e31827f2481] [Medline: 23417218]
41. Maternity, paternity and parental allowances. Kela. 2021. URL: https://www.kela.fi/web/en/parental-allowances [accessed 2022-01-01]
42. Niela-Vilén H, Auxier J, Ekholm E, Sarhaddi F, Asgari Mehrabadi M, Mahmoudzadeh A, et al. Pregnant women's daily patterns of well-being before and during the COVID-19 pandemic in Finland: longitudinal monitoring through smartwatch technology. PLoS One 2021 Feb 3;16(2):e0246494 [FREE Full text] [doi: 10.1371/journal.pone.0246494] [Medline: 33534854]
43. Yeragani VK, Sobolewski E, Kay J, Jampala VC, Igel G. Effect of age on long-term heart rate variability. Cardiovasc Res 1997 Jul;35(1):35-42. [doi: 10.1016/s0008-6363(97)00107-7] [Medline: 9302345]
44. Kim HG, Cheon EJ, Bai DS, Lee YH, Koo BH. Stress and heart rate variability: a meta-analysis and review of the literature. Psychiatry Investig 2018 Mar;15(3):235-245 [FREE Full text] [doi: 10.30773/pi.2017.08.17] [Medline: 29486547]
45. Nguyen PT, Rahman MS, Le PM, Van Nguyen H, Vu KD, Nguyen HL, et al. Trends in, projections of, and inequalities in reproductive, maternal, newborn and child health service coverage in Vietnam 2000-2030: a Bayesian analysis at national and sub-national levels. Lancet Reg Health West Pac 2021 Jul 30;15:100230 [FREE Full text] [doi: 10.1016/j.lanwpc.2021.100230] [Medline: 34528011]

Abbreviations

AVNN: average normal interbeat intervals
bpm: beats per minute
ECG: electrocardiogram
HF: high-frequency power
HLM: hierarchical linear mixed
HR: heart rate
HRV: heart rate variability
IBI: interbeat interval
IoT: Internet of Things
LF: low-frequency power
nHF: normalized high-frequency power
nLF: normalized low-frequency power
nRMSSD: normalized root mean square of the successive difference of normal interbeat intervals
nSDNN: normalized SD of normal interbeat intervals
PPG: photoplethysmogram
RMSSD: root mean square of the successive difference of normal interbeat intervals
SDNN: SD of normal interbeat intervals

Edited by L Buis; submitted 09.09.21; peer-reviewed by P Nguyen, D Filos; comments to author 17.11.21; revised version received 12.01.22; accepted 03.03.22; published 03.06.22.

Please cite as:
Sarhaddi F, Azimi I, Axelin A, Niela-Vilen H, Liljeberg P, Rahmani AM
Trends in Heart Rate and Heart Rate Variability During Pregnancy and the 3-Month Postpartum Period: Continuous Monitoring in a Free-living Context
JMIR Mhealth Uhealth 2022;10(6):e33458
URL: https://mhealth.jmir.org/2022/6/e33458
doi:10.2196/33458
PMID:35657667

Original Paper

Conflicting Aims and Values in the Application of Smart Sensors in Geriatric Rehabilitation: Ethical Analysis

Christopher Predel[1], MD; Cristian Timmermann[1], DPhil; Frank Ursin[1], DPhil; Marcin Orzechowski[1], Dr rer soz; Timo Ropinski[2], Prof Dr; Florian Steger[1], Prof Dr

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

[2]Visual Computing Group, Ulm University, Ulm, Germany

Corresponding Author:
Christopher Predel, MD
Institute of the History, Philosophy and Ethics of Medicine
Ulm University
Parkstraße 11
Ulm, 89073
Germany
Phone: 49 228257273
Email: christopher.predel@uni-ulm.de

Abstract

Background: Smart sensors have been developed as diagnostic tools for rehabilitation to cover an increasing number of geriatric patients. They promise to enable an objective assessment of complex movement patterns.

Objective: This research aimed to identify and analyze the conflicting ethical values associated with smart sensors in geriatric rehabilitation and provide ethical guidance on the best use of smart sensors to all stakeholders, including technology developers, health professionals, patients, and health authorities.

Methods: On the basis of a systematic literature search of the scientific databases PubMed and ScienceDirect, we conducted a qualitative document analysis to identify evidence-based practical implications of ethical relevance. We included 33 articles in the analysis. The practical implications were extracted inductively. Finally, we carried out an ethical analysis based on the 4 principles of biomedical ethics: autonomy, beneficence, nonmaleficence, and justice. The results are reported in categories based on these 4 principles.

Results: We identified 8 conflicting aims for using smart sensors. Gains in autonomy come at the cost of patient privacy. Smart sensors at home increase the independence of patients but may reduce social interactions. Independent measurements performed by patients may result in lower diagnostic accuracy. Although smart sensors could provide cost-effective and high-quality diagnostics for most patients, minorities could end up with suboptimal treatment owing to their underrepresentation in training data and studies. This could lead to algorithmic biases that would not be recognized by medical professionals when treating patients.

Conclusions: The application of smart sensors has the potential to improve the rehabilitation of geriatric patients in several ways. It is important that patients do not have to choose between autonomy and privacy and are well informed about the insights that can be gained from the data. Smart sensors should support and not replace interactions with medical professionals. Patients and medical professionals should be educated about the correct application and the limitations of smart sensors. Smart sensors should include an adequate representation of minorities in their training data and should be covered by health insurance to guarantee fair access.

(JMIR Mhealth Uhealth 2022;10(6):e32910) doi:10.2196/32910

KEYWORDS

personal data; wearable; older adults; autonomy; rehabilitation; smart sensor; machine learning; ethics; access to health care; justice

Introduction

Regular physical activity reduces the risk of many chronic diseases and can significantly contribute to rehabilitation. Geriatric patients are often affected by reduced exercise capacity, which leads to mobility restrictions and dependence on support in daily life [1]. Diagnostic methods can be used to assess physical activity levels for enhancing rehabilitation. These include patient-reported outcomes and clinical gait analyses. A limitation of the methods currently in use is that the delivered data are often difficult to objectify [2]. To overcome this limitation, technology developers and physicians have begun to use smart sensors [3].

Smart sensors combine the measurement and analysis of data. They can collect a wide range of data and can be used in different application areas [4]. In this analysis, we focus on the ethical evaluation of smart sensors that use inertial sensors and machine learning algorithms to record and analyze complex movement patterns. For this purpose, patients receive wearable inertial sensors that record the acceleration in space. Using machine learning techniques, these inertial data can be assigned to complex movement patterns, such as standing up from a chair, opening a door, or even falling. Thus, it is possible to record the activity patterns of patients and quantify their daily activity [5]. In this manner, clinicians can objectively assess patients' daily physical activity and identify their treatment needs. Care and rehabilitation measures can be individually adapted, and treatment progress can be documented [6]. Rehabilitation of geriatric patients using smart sensor technology has the potential to increase the quality of life for many patients. However, the recording of such data monitors all daily activities can be negatively associated with patient surveillance.

The high vulnerability of geriatric patients and the special characteristics of machine learning algorithms also raise ethical challenges, which will be discussed in this paper. We concentrate our research on the following question: What are the ethical challenges of using smart sensors and how can they be minimized? Our goal is to identify and analyze the different ethical values associated with smart sensors and their potential conflicts, and based on this ethical analysis, provide guidance to all stakeholders, including technology developers, health professionals, patients, and health authorities.

Methods

This research is an ethical analysis that aims to examine the ethical challenges associated with smart sensors in geriatric rehabilitation.

Systematic Literature Search

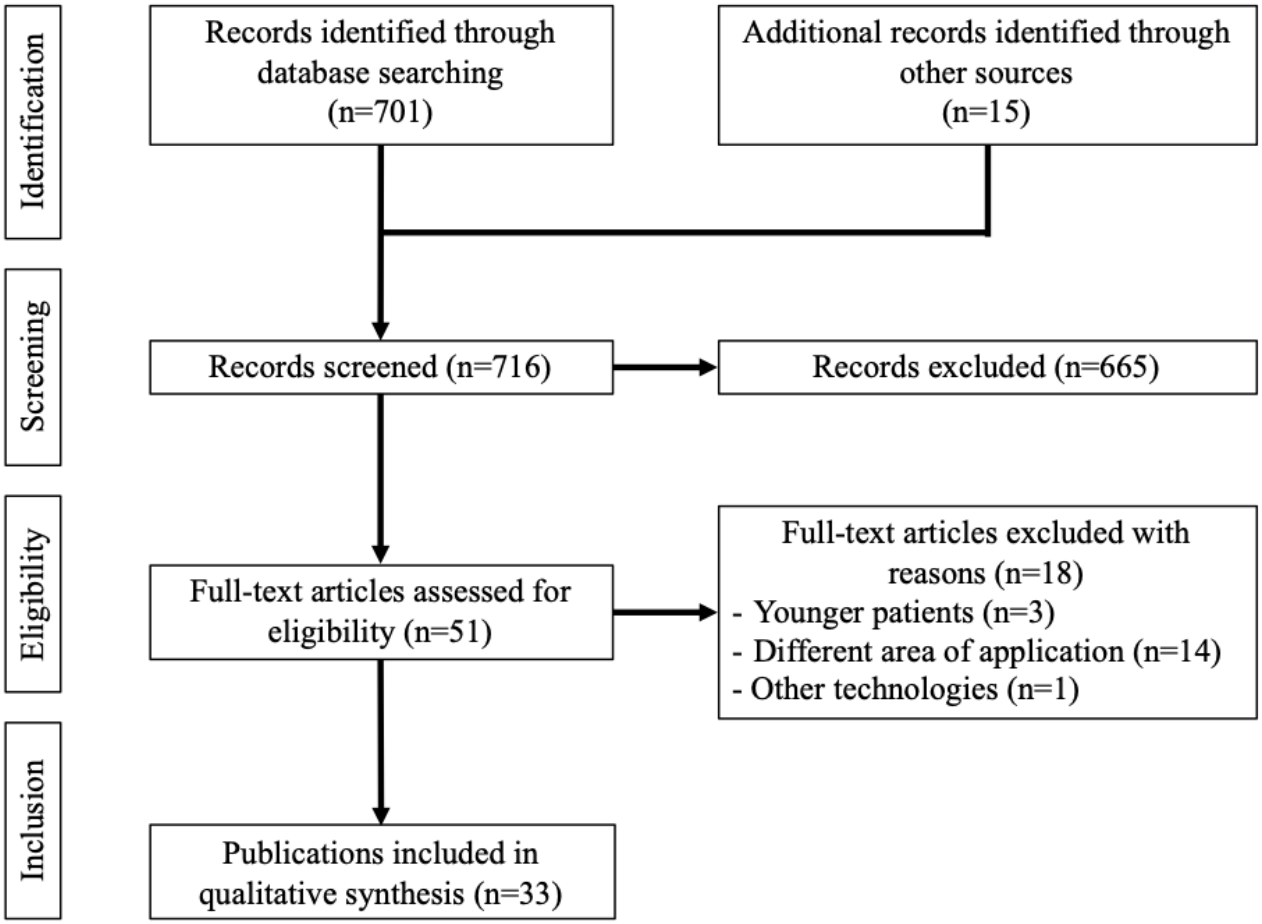
First, the literature on smart sensors in geriatric rehabilitation was identified through a systematic literature search. We then inductively extracted evidence-based practical implications of ethical relevance through qualitative document analysis. PubMed and ScienceDirect databases were used to identify published literature between January 2000 and November 2020. The search was supplemented by using Google Scholar. The literature search was carried out using the following steps: first, identification and definition of the research question and creation of a search algorithm; second, identification of relevant studies; third, selection of studies; and fourth, reporting of the results in an ethical analysis based on the principle-oriented approach of Beauchamp and Childress [7]. Therefore, we combined 2 research methods that are frequently used to assess the ethical issues of new developments in medical practice: a systematic review of all ethical aspects and a systematic review of all ethical values [8,9].

As smart sensors are a novel technology, common synonyms and related terms have been used to avoid missing relevant literature. The search algorithm combined the keywords *smart sensor*, *wearable electronic devices*, *wearable*, *intelligent assistive technology* and *internet of things* with the keywords *geriatric*, *elderly*, *rehabilitation*, or *dementia* and *ethics*, *privacy*, *empowerment*, *harm*, *caregiver*, *discrimination*, *informed consent* or *autonomy* in the titles and abstracts of articles.

Owing to the limited number of eligible ethical analyses, articles on the use of sensors in the care of older adults, in general, were also included. The results of these articles were translated by analogy to the application of rehabilitation. Articles that discussed only the implementation, development, or technical specifications of sensor technologies or algorithms were excluded. No restrictions on article type were imposed.

The search algorithm yielded 701 results (Figure 1). Additional 15 articles were identified through hand search using Google Scholar. After removing duplicates and screening the titles and abstracts, 51 articles were considered eligible. After reviewing the full text, 18 articles were excluded because they did not meet the inclusion criteria. The excluded articles focused on younger patients, analyzed different purposes of the application such as sports or lifestyle, or analyzed other technologies, such as robots.

Figure 1. Flowchart of the systematic literature search to identify evidence-based practical implications of applying smart sensors in geriatric rehabilitation resembling the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement.



A total of 33 articles were included in this ethical analysis. First, the content of the articles was screened for key information of ethical significance. The content of these articles contains evidence-based practical implications for the application of smart sensors in geriatric rehabilitation.

Thematic Analysis

Second, a thematic analysis was performed. This is a qualitative approach for identifying, analyzing, and reporting common patterns or themes in narratives or text materials. Articles were explored for recurring themes with a focus on the different values and aims of smart sensors in geriatric rehabilitation [10,11].

Ethical Analysis

Third, the identified practical implications of ethical relevance were grouped, and an ethical analysis was conducted using the principle-oriented approach of Beauchamp and Childress [7]. If an ethical issue could be examined under more than one ethical principle, we opted to report the issue under the principle that was better suited to highlight ethical conflicts. For reducing biases and omissions, the included articles were critically examined by at least two authors, as recommended for systematic reviews of normative literature [12,13]. We pooled the main ethical issues together after an exchange between the authors in dichotomous pairs of conflicting aims and values. In the following sections, we propose an assessment tool for the ethical evaluation of smart sensors in geriatric rehabilitation. With our tool, physicians, along with their patients, will be able to assess which values are more important to them in each individual case and then weigh the different values against each other.

Results

In this section, we report the ethical challenges identified in the systematic literature search, grouping them under the 4 principles of biomedical ethics: autonomy, beneficence, nonmaleficence, and justice.

Autonomy

Respect for autonomy is a fundamental principle of biomedical ethics and requires ensuring that the patient's will is respected, unless it is in direct conflict with other fundamental values and professional duties. It includes the negative obligation to not constrain a patient's actions unnecessarily and the positive obligation to disclose information that fosters decision-making. Measures that empower patients tend to increase their autonomy, whereas interventions that directly restrict their liberties or make them hesitant to act freely, restrict patients' autonomy.

The use of smart sensors in rehabilitation can empower patients by increasing their proactive participation in diagnostics and allowing them an independent life at home. Patients who were asked about the use of wearables in rehabilitation indicated that

they expected to be empowered by this technology to manage their own health conditions more effectively [14]. Continuous feedback on the progress of the rehabilitation can motivate patients to become physically active and continue therapy [14-16]. Furthermore, it could provide patients with a deeper understanding of their illness and physical condition [17]. Efficient rehabilitation, aided by smart sensors, can reduce the need for long-term care. Rehabilitation can be supplemented by fall detection and home monitoring, enabling patients to stay at home independently for longer [18,19]. In addition, 58% of patients using fall detectors had improved independence and 72% felt more confident [20]. Through their proactive participation in health management and the possibility of living at home independently for a longer time, patients' autonomy is increased by the use of smart sensors.

Privacy

Privacy can be defined as an interest, or even as a right, to be free from intrusion in personal matters, unless major public interests justify such an invasion [21]. When using smart sensors, the protection of privacy requires a person to be left alone when asked and not be monitored without expressed wishes. In contrast, data privacy is concerned with the sensitive handling of data, including their access and use by third parties [22]. Privacy concerns are one of the biggest hurdles for patients in the application of supportive technologies [23]. In a study, Canadian stakeholders were interviewed regarding the challenges of active assisted living technologies. In 30% of the mentions, privacy and security were identified as primary issues [24]. Monitoring patients' day-to-day activities is highly intrusive. The feeling of being constantly monitored and ubiquitous medical diagnostics can lead to stress and anxiety and may compel patients to adapt their behavior. The evaluation of all daily activities and a desire to achieve good measurement results can lead to excessive physical activity. In a study in which the daily physical activity of patients with chronic obstructive pulmonary disease was measured using sensors, it was shown that participants had a 26% higher activity than the average during the first few days monitored with sensor technology [25]. In the context of rehabilitation of geriatric patients, this may lead to stress and overload symptoms. In consequence, injuries and falls can occur more often.

Patients' perceptions of privacy loss are significantly influenced by the intrusiveness of the technology used [18,26]. Owing to their low-threshold use, smart sensors offer the potential to minimize the feeling of surveillance through a low degree of intrusiveness and by only collecting data related to preselected complex movement patterns. Studies have shown that patients using smart sensors do not feel violated in their privacy [27]. Patients prefer sensors that can only monitor whether they are active and do not identify specific activities [28]. In most cases, it is not clear whether, to what extent and by whom, the gathered data could be analyzed to conclude information about patients that was not willingly shared by them.

Patients must be informed of the conclusions drawn from the data. From movement data, it is possible, for example, to analyze how often patients use the bathroom, whether they drink alcohol, or whether they are sexually active. It must be discussed with the patient which activities should and could be tracked. Patients should be trained to switch off or dismount sensors when privacy is desired so that they are free to undertake the activities they value and do not have to make unnecessary sacrifices to maintain an image of themselves that they are comfortable sharing with the medical team. As nonmaleficence demands not depriving people from a good they value, loss of privacy can also be seen as a form of harm [29].

Shared Decision-making

When adequately introduced, the use of smart sensors can improve the *patient-medical professional* relationship and increase autonomy by strengthening the patient's role as an equal partner. Medical professionals and patients can make therapeutic decisions together, based on data collected by the patient [16,30,31].

Empirical studies have assessed the impact of smart sensors on patient-physician relationships. Patients were asked whether they expected a change in their relationship with their medical professional through wearable technology during rehabilitation. They stated that they expected an improvement in communication and a more patient-centered consultation due to the improved and objective data gathered on their activities [14]. Patients using smart sensors expressed that they were well informed and that decision-making between medical professionals and patients could be improved [32]. Furthermore, the diagnostic process is no longer limited to a visit to the medical practice or hospital; it also takes place beyond these settings. Thus, patients can receive medical support in everyday life [33]. Smart sensor technology used at home can be designed to facilitate contact with medical professionals [34,35].

Beneficence

Overview and General Aspects

The principle of beneficence dictates the orientation of health professionals' actions toward the well-being of patients. This demands that health professionals make use of both their professional and interpersonal skills to improve the situation of patients, particularly in helping them to fulfill their wish to live in their own homes, while ensuring that such choices do not come at the cost of losing all types of bonds with them.

An advantage of smart sensors is the possibility of independent home monitoring. Long-term home monitoring can provide objective movement data on patients' everyday life. This can increase the well-being of patients by allowing them to remain in the comfort of their own homes, but it can also reduce the number of social contacts [27,36]. In addition, sensors can contribute to patient safety by extending the monitoring phase after surgery or by identifying patients at a risk of adverse events.

Objective Assessment of Daily Activities

Previously, therapy requirements and progress have been determined using gait analyses or patient-reported outcomes. These have high inter- and intraobserver variability and are mostly carried out in a clinical setting [37]. In contrast, sensor technology offers the possibility of objective long-term home monitoring. This has the advantage that the complex movement

patterns of patients can be analyzed in everyday situations and over a longer period [38]. Decisions for or against a rehabilitative measure can therefore be made from a broad database. Patients can receive therapy adapted to their everyday life [16]. Thus, autonomy and independence can be promoted in everyday life to benefit patients.

Extended Monitoring and Injury Prevention

The use of smart sensors can lead to more individualized therapies for geriatric patients in their homes. Geriatric patients have a high acceptance of sensory technology for long-term home monitoring during rehabilitation [28]. After surgery, patients often remain in the ward for several days for monitoring. Sensor technology can significantly extend monitoring time without the need to keep patients hospitalized. Thus, treatment needs, which only become clear in the patient's everyday life after discharge, can be identified. Patients benefit from greater security without having to spend more time in the hospital. Owing to the low-threshold use of smart sensors, opportunities for screening and prevention have expanded. People who are expected to need treatment in the future because of hospital stays, comorbidities, or old age can wear sensors in their everyday life. If conspicuous movement patterns appear, practitioners can be informed, enabling them to assess an intervention or rehabilitation need. Thus, the user can benefit from preventive intervention [3]. Furthermore, sensor technology can be used by risk groups to identify and prevent critical events, such as falls [39]. For many patients, an increased sense of security is one of the main reasons for using sensor technology [40,41]. A total of 85% of patients who used a fall detector stated that it improved their safety [20].

The large amount of data collected by smart sensors can be used by machine learning algorithms to detect different anomalies and then take early steps to address health threats. If a patient goes to the bathroom more often than usual, it could be a sign of urinary tract infection or diabetes mellitus. A decrease in the number of outdoor activities could be a result of depression. It must be determined which activities the sensor technology should record and whether findings must be interpreted as relevant for rehabilitation.

Nonmaleficence

The principle of nonmaleficence indicates that new medical technologies should not disadvantage or harm patients through medical intervention or even diagnosis. The biggest threats to using smart sensors in geriatric rehabilitation are the misuse of patients' private data and the uncritical acceptance of data provided by the sensors. A major threat to patients is the misuse of data by unauthorized persons. Cyberattacks can steal data from various devices and servers. Owing to the interconnectivity between smart sensors and digital health records, as well as the multiple users and use outside of protected hospital networks, smart sensors represent vulnerable targets for cyberattacks [42].

Accuracy

In the detection of complex movement patterns, inaccurate activity detections can occur and cause harm to patients. Algorithms may not recognize or they may misclassify movements [19]. Incorrectly classified events can lead to an overestimation of patient's health. Conversely, the need for rehabilitation or lack of therapeutic success can be overlooked [17]. Therefore, uncritical acceptance of movement data by medical professionals poses a risk to patients. Sensors can support the medical professional's subjective assessment of care needs with objective data, but cannot replace a complete examination [43]. By increasing the autonomy of patients, there has also been a shift in the roles of patients and medical professionals. The patient is the one who has to apply the sensor technology. As a result, the expectation is placed on the patient to provide high-quality data. Therefore, patients gain more responsibility in the diagnostic process. This could lead to more autonomy but could also jeopardize data quality [44].

Missing Holistic Assessment

By reducing direct contact with medical professionals and relying more on smart sensors, there is a risk of patients being reduced to the data collected [17]. Social contact with medical professionals is an essential component of therapy. Collecting data on only one physiological parameter, such as movement patterns, does not provide a holistic assessment of health conditions and the rehabilitation process. A holistic assessment can only be discerned through direct interaction with health care professionals [27,34]. Successful treatment requires contact with a medical professional who communicates the results of a diagnosis with empathy and is aware of the patient's circumstances [30]. The feeling of being monitored can reduce the trust between patients and medical professionals and the acceptance of sensor technology. Moreover, patients may overestimate the accuracy and potential of smart sensors [19].

Justice

Overview and General Aspects

The principle of justice refers to 2 distinct principles: first, that like cases be treated alike and second, to a fair, equitable, and appropriate distribution of health care in society. This demands that every patient should have adequate access to essential health care, regardless of gender, ethnicity, sexual orientation, religion, age, or socioeconomic status [7]. Smart sensors are expensive and can therefore lead to discrimination on the basis of socioeconomic differences. Owing to the dependence on the accuracy of the training data, algorithmic analyses could lead to a discrimination against minorities that are underrepresented in the training data. Geriatric patients who have less experience with technical tools can be at a disadvantage. Conversely, patients living in underserved regions may benefit from the use of sensors in combination with telemedicine. In addition, the success of rehabilitation measures aided by smart sensors depends on the capability of users to use digital technologies, or more broadly, their digital literacy.

Socioeconomic Differences

High prices during early technology adoption lead to inequalities in access to personalized rehabilitation. For many patients, the acceptance and adoption of sensors depends on their cost [45]. The use of wearable sensors such as smartwatches shows major demographic and socioeconomic differences. Mainly young, wealthy people buy smartwatches [23]. An additional negative consequence is that smart sensors are optimized on these early

adopters, basing the algorithms and the design of the software and hardware on a subpopulation that does not reflect the diversity of the population with rehabilitation needs. New developments that do not solely rely on external systems and are adapted for the geriatric population could overcome this limitation. If sensors are not covered by health insurance and must be purchased by the patients themselves, there will be major inequalities in the medical care of the population [46].

Discrimination of Vulnerable Groups

The diagnostic accuracy of smart sensors and the algorithms used by them depends on the training data. Thus, there are differences in accuracy depending on the population group. Population groups that are underrepresented in the training data do not benefit from a high algorithmic output accuracy. They must adapt to the standard defined by the training data even if their movement patterns are normal for their group [17,47]. Furthermore, the movement patterns of men and women differ in some aspects. An algorithm trained using male movement data has a higher output accuracy for men than for women. Similar conclusions can be drawn for other population groups, such as older adults. Studies have shown that it is possible to predict the gender and age of participants with inertial data from gait analysis [26].

There is also a risk of disparity between age groups. The use and function of technical devices are difficult to understand for many older adults. The application of smart sensors in the context of geriatric rehabilitation requires extensive training and education of patients so that they can learn the limitations and correct application of sensor technology and thus benefit from its advantages [46]. Monitoring technologies can cause feelings of stigma and frailty in geriatric patients [27]. Their use can be seen by patients as an admission of frailty and illness to themselves and the social environment [48]. Wearing sensors in public can reveal illnesses or disabilities to strangers [27,49]. In order to mitigate this, smart sensors can be integrated into clothes or smart watches [50]. By giving patients the opportunity to choose between different types of application, the feeling of stigma can be actively reduced.

Increasing Numbers of Patients Can Be Treated

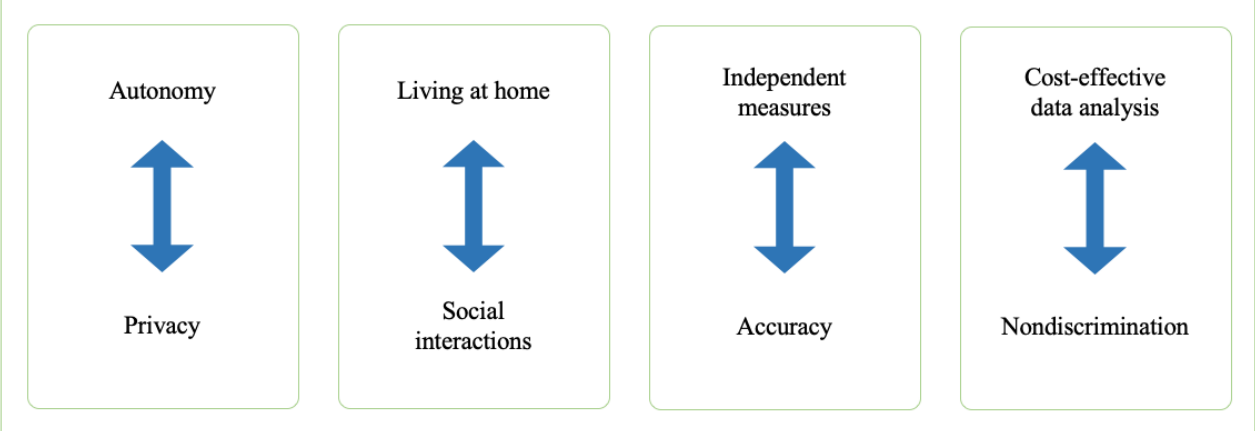
Smart sensors have the potential to provide high-quality care to each patient. The quality of human-influenced treatment depends heavily on the experiences, prejudices, and daily constitution of medical professionals. Smart sensors developed and evaluated in congruence with ethical principles offer the possibility of consistently delivering high-quality treatment [51]. Owing to automated data collection and processing, smart sensors offer the possibility of treating more patients at a consistent and even higher quality of care. In many places, there is a supply gap between urban and rural areas in specialized medical care. By using smart sensors in combination with telemedicine, patients in underserved regions can be connected to medical specialists [3]. As previously discussed, this requires extensive training, which not all patients, especially geriatric patients, can follow. Furthermore, fair access to new promising technologies, such as smart sensors, must be guaranteed in rural areas.

Discussion

Principal Findings: Conflicting Aims and Values

Our ethical analysis showed that the rehabilitation of geriatric patients can generally be improved using smart sensors. However, we found conflicting values and aims that doctors and patients must consider when using smart sensors for rehabilitation. The use of smart sensors involves 4 pairs of conflicting ethical values and aims, which patients should sufficiently understand to provide informed consent and maintain compliance with optimal use (Figure 2).

Figure 2. Conflicting aims and values.



Proactive participation in diagnostics and gaining independence can increase patient autonomy. However, gains in autonomy come at the cost of privacy. Owing to the continuous monitoring of patients' daily activities, privacy can be violated if sensors are too intrusive, and patients have no control over their data. Moreover, when patients are aware that they are being surveilled, they may refrain from doing certain things that they value.

In contrast to smart sensors in dementia care, sensors that are used in rehabilitation are not intended to be used for surveillance, but for promoting autonomy by assisting rehabilitation measures. Increased autonomy and the benefits

of home monitoring conflict with protecting patients' privacy. The treatment team receives sensitive information using sensor technology in everyday life. The use of wearable technologies carries the risk of increasing the intended or unintended disclosure of sensitive health information [52]. This information is not consciously shared by patients with a specific health professional but is automatically collected by a technical instrument. It might not always be clear to patients who has access to this information and what the data reveal. To guarantee confidentiality of the information collected by sensors, authorized recipients must be specifically defined.

It is important that the patient be educated about the extent of the invasion of their privacy. Therefore, medical professionals must understand what conclusions can be drawn from the data in addition to the daily activity patterns. It is crucial to keep in mind that future developments could allow further data analysis and thus reveal unforeseen information, which could extend the invasion of patient privacy. Smart sensors are an attractive target for cyberattacks, because they collect valuable data and are often used in unprotected private settings. To protect the privacy of patients, it is important that service providers protect the data from unauthorized access and misuse. Regular secure backups, anonymization of the data, and limiting remotely accessible data can reduce the risk of data theft. Patients must be adequately informed and educated about this risk, ways to reduce it, and how they can avoid being monitored when privacy is desired [42].

Studies have shown that most patients do not feel that their privacy is violated by the use of smart sensors and are willing to give up some of their privacy for increased autonomy [28,41,53]. Depending on the amount of autonomy gained and the degree of invasion of privacy, there is a different level of willingness to use this technology. Older people who have an increased risk of falls or who would benefit from rehabilitative measures could consent to the invasion of their privacy by motion sensors in exchange for increased safety and autonomy [37,47]. In contrast, less vulnerable patients may have fewer reasons to allow wider intrusion in their personal life. Overall, patients need to weigh the autonomy gained with the use of smart sensors against eventual losses of autonomy by feeling compelled to adapt their behavior when monitored.

Independent measurements are the principal reason for using smart sensors at home and for monitoring daily activities; however, if patients are not sufficiently trained in the use of the sensors, it can lead to decreased accuracy of the data. Independent measures can increase patient autonomy and provide the opportunity to monitor daily activity patterns; however, they come at the cost of a decreased number of social interactions with medical professionals and reduced accuracy. Independent measurements provide the opportunity to live longer at home and generate objective data that represent daily activity patterns, but they could reduce the number of social interactions with medical professionals.

It is important that the sensor technology and underlying algorithm be supportive and not replace the diagnostic process. Before deciding for or against an intervention, the treatment team should have direct contact with the patient [54]. Sensors can support medical professionals' subjective assessments of care needs using objective data [43]. The treatment team should critically question and contextualize the algorithmic output at any time [17]. As smart sensors reduce the number of social contacts with medical professionals, it is important to keep in mind that solitude is one of the largest welfare and mental health issues among older adults [55]. Although medical treatment may be the only social activity of a significant number of older adults, it should be noted that such interactions do not solve the problem of solitude. Better alternatives outside the therapeutic context should be offered for public mental health.

To improve the accuracy of the sensor technology, developers need to work on the accuracy of smart sensors if they are used for monitoring patients at risk. Medical professionals need to be aware that some measures that require high precision may need to be carried out under their direct supervision and that there are limits on what can be accurately measured outside clinical settings. Training should be given on the correct application of the sensors to empower the patient to increase the accuracy of the measurements.

To justify the use of public health resources, it is necessary to prove the increased effectiveness of sensor technology compared with conventional methods. A cost-effectiveness calculation of the use of smart sensors needs to fully recognize the multiple advantages of increased mobility for older adults' well-being. In view of the long-term health benefits of increased mobility, access to smart sensors for rehabilitation should be independent of the patient's socioeconomic status. To guarantee fair distribution, sensor technology should be prescribed by a physician and covered by health insurance. To ensure patient participation in areas with limited access, the technology should be designed such that it can be used independently or at least with the easy assistance of family members. Specialists can be contacted during anomalies [34]. We conclude that smart sensors can provide high-quality, low-cost measurement tools for many patients. However, because algorithms are seldom developed and tested for diverse populations, minorities may be at a disadvantage.

With regard to the principle of social justice, the provision of modern health care appliances for patients, such as smart sensors, requires that they are able to efficiently use them in their daily life. Smart sensors can enhance access to health care for underserved populations. However, here, as in the case of other digital instruments in health care, the opportunities provided by smart sensors are subjected to adequate use and can result in significant inequalities with respect to who can use and benefit from them [56]. The foremost is the ability to understand and use digital technologies, *digital literacy*. This ability is heterogenic and conditioned by several components; for example, skills, resources, and motivations. It has been observed that the level of literacy in the use of digital technologies is associated with social attributes of patients, such as age, level of education, health literacy in general, language barriers, immigration status, and urban or rural residence [57]. Older adult users face additional barriers when using digital technologies [58]. Extensive training and education are required regarding the use of smart sensors. Deficits in trust in digital health instruments, lack of previous experience with similar

appliances, low levels of education, or language barriers can significantly impede this process.

Smart sensors have been used in geriatric rehabilitation. It has been shown that sensors can support the rehabilitation process by providing objective monitoring of a patient's activity level [6,59]. Thus, medical professionals can define and examine rehabilitation targets along with patients to track the process. By using sensors, it is possible to compare the individual progress of a patient with the expected average progress of other patients with similar comorbidities. Activity levels can be tracked outside the therapy session. The data could be used to justify the extension of rehabilitation measures to insurance companies [6,59]. The current implementations have already addressed some of the ethical challenges mentioned in this paper, but they were used in a hospital setting. Patients were always able to communicate problems or discomfort with the sensors to medical professionals. To decrease the feeling of surveillance, the sensors were located on the lower back of the patients [6].

Recommendations

Our principal recommendation is to consider multiple factors affecting digital literacy in the process of patient education to facilitate the effective use of smart sensors. Second, patients should not have to decide between autonomy and privacy. Developers should aim at providing solutions that promote patient autonomy while also ensuring privacy by collecting minimal amounts of data necessary to operate effectively. The standard for the ethical implementation of smart sensors should follow four prerequisites: (1) smart sensors can be activated and deactivated by the patient, (2) smart sensors are not visible to the public, (3) smart sensors only collect activity data over which a patient has control, and (4) they collect the minimal amount of data needed to allow an accurate diagnosis. In some cases, we may observe that patients refuse to sacrifice their privacy for increased autonomy. In such cases, it must be evaluated together with patients whether and to what extent this intrusion into privacy needs to be tolerated, how it can be minimized, and how great the actual benefit of sensors is for the patient in comparison with alternative treatment options.

Further recommendations for developers, patient education, health professionals, and health authorities are summarized in Textbox 1.

Textbox 1. Recommendations for developers, medical professionals, and health authorities.

Developers

- Authorized recipients that have access to specific data must be defined.
- Data need to be protected from unauthorized access and misuse.
- Smart sensors should be activated and deactivated by the patient.
- Smart sensors should not be visible to the public.
- Smart sensors should only collect activity data over which a patient has control.
- Minimal amount of data needed to allow an accurate diagnosis should be collected.
- In order to ensure patient participation in areas with limited access, the technology should be designed so that it can be used independently, or at least easily, with the assistance of family members.
- Contact with specialists in the event of anomalies should be facilitated.

Patient education

- Education of the patient about the extent of invasion of privacy and the conclusions that can be drawn from the data must be done.
- Training should be given on the correct application of the sensors to empower the patient to increase the accuracy of measurements.

Medical professionals

- Smart sensors should augment and not replace the diagnostic process. The treatment team should have direct contact to the patient.
- Algorithmic outputs should be contextualized and questioned critically.
- Medical professionals should be aware of the limits and accuracy of smart sensors.

Health authorities

- It is necessary to prove an increased effectiveness of sensor technology compared with conventional methods to justify the use of public health resources. A cost-effectiveness calculation of the use of smart sensors needs to fully recognize the multiple advantages that increased mobility has for older adults' well-being.
- To guarantee a fair distribution, sensor technology should be prescribed by a physician and covered by health insurance.

Limitations and Comparison With Prior Work

There are already several articles that analyzed the ethical challenges of smart wearable sensors, but no article focused on smart sensors for geriatric rehabilitation [18,31]. Much of the current literature primarily discusses the ethical challenges of intelligent assistive technologies for monitoring geriatric patients, particularly in dementia care [18,60]. There are also articles that discuss issues with smart sensors used for activity and mobility monitoring. These articles focus on healthy or younger participants and rarely discuss the issues of smart sensors used by geriatric patients in rehabilitation [61,62]. Some articles discuss the use of other technologies, such as telemedicine or apps for self-management and tracking in rehabilitation [63,64]. However, these articles do not analyze the specific ethical issues associated with tools that are based on machine learning algorithms.

A limitation of this study is that it did not examine the subjective perceptions of the main stakeholders. Empirical ethical studies in the field of smart sensors are insufficient. Further work is needed to investigate the ethical insights of health professionals using smart sensors and to study the experiences of patients who use such sensors.

Conclusions

Smart sensors offer an opportunity for the objective assessment of complex movement patterns and rehabilitation progress. Medical professionals must consider and address multiple conflicting ethical aims. One conflict in aims is that gains in autonomy often come at the cost of patient privacy. It is important that patients are educated on the insights that the collected data reveal and do not have to decide between autonomy and privacy. Furthermore, smart sensors should not replace but instead promote interaction with medical professionals. As smart sensors are complex and novel tools, medical professionals and patients should be educated on their correct applications and their limitations. Sensors should be covered by insurance to guarantee equal access to health care.

Acknowledgments

This paper is part of the *Identification of Complex Activity Patterns through Smart Sensors in Geriatric Rehabilitation* (AktiSmart-KI) project funded by the Bundesministerium für Gesundheit (BMG).

Conflicts of Interest

None declared.

References

1. Motl RW, McAuley E. Physical activity, disability, and quality of life in older adults. Phys Med Rehabil Clin N Am 2010 May;21(2):299-308. [doi: 10.1016/j.pmr.2009.12.006] [Medline: 20494278]
2. Polhemus AM, Bergquist R, Bosch de Basea M, Brittain G, Buttery SC, Chynkiamis N, et al. Walking-related digital mobility outcomes as clinical trial endpoint measures: protocol for a scoping review. BMJ Open 2020 Jul 19;10(7):e038704 [FREE Full text] [doi: 10.1136/bmjopen-2020-038704] [Medline: 32690539]
3. Patel S, Park H, Bonato P, Chan L, Rodgers M. A review of wearable sensors and systems with application in rehabilitation. J Neuroeng Rehabil 2012 Apr 20;9:21 [FREE Full text] [doi: 10.1186/1743-0003-9-21] [Medline: 22520559]
4. Chan M, Estève D, Fourniols J, Escriba C, Campo E. Smart wearable systems: current status and future challenges. Artif Intell Med 2012 Nov;56(3):137-156. [doi: 10.1016/j.artmed.2012.09.003] [Medline: 23122689]
5. Penteridis L, D'Onofrio G, Sancarlo D, Giuliani F, Ricciardi F, Cavallo F, et al. Robotic and sensor technologies for mobility in older people. Rejuvenation Res 2017 Oct;20(5):401-410. [doi: 10.1089/rej.2017.1965] [Medline: 28482748]
6. Klenk J, Wekenmann S, Schwickert L, Lindemann U, Becker C, Rapp K. Change of objectively-measured physical activity during geriatric rehabilitation. Sensors (Basel) 2019 Dec 11;19(24):5451 [FREE Full text] [doi: 10.3390/s19245451] [Medline: 31835673]
7. Principles of Biomedical Ethics. Oxfordshire, England, UK: Oxford University Press; 1979.
8. Mertz M. How to tackle the conundrum of quality appraisal in systematic reviews of normative literature/information? Analysing the problems of three possible strategies (translation of a German paper). BMC Med Ethics 2019 Nov 14;20(1):81 [FREE Full text] [doi: 10.1186/s12910-019-0423-5] [Medline: 31727134]
9. Mertz M, Strech D, Kahrass H. What methods do reviews of normative ethics literature use for search, selection, analysis, and synthesis? In-depth results from a systematic review of reviews. Syst Rev 2017 Dec 19;6(1):261 [FREE Full text] [doi: 10.1186/s13643-017-0661-x] [Medline: 29258598]
10. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: implications for conducting a qualitative descriptive study. Nurs Health Sci 2013 Sep;15(3):398-405. [doi: 10.1111/nhs.12048] [Medline: 23480423]
11. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
12. Kahrass H, Borry P, Gastmans C, Ives J, van der Graaf R, Strech D, et al. PRISMA-Ethics – Reporting Guideline for Systematic Reviews on Ethics Literature: development, explanations and examples. OSF 2021 (forthcoming) [FREE Full text]

13. Strech D, Sofaer N. How to write a systematic review of reasons. J Med Ethics 2012 Feb;38(2):121-126 [FREE Full text] [doi: 10.1136/medethics-2011-100096] [Medline: 22080465]
14. Belsi A, Papi E, McGregor AH. Impact of wearable technology on psychosocial factors of osteoarthritis management: a qualitative study. BMJ Open 2016 Feb 03;6(2):e010064 [FREE Full text] [doi: 10.1136/bmjopen-2015-010064] [Medline: 26842273]
15. Barton AJ. Wearable technology to enhance remote monitoring and self-management. Clin Nurse Spec 2015;29(1):6-7. [doi: 10.1097/NUR.0000000000000088] [Medline: 25469433]
16. Appelboom G, Camacho E, Abraham ME, Bruce SS, Dumont EL, Zacharia BE, et al. Smart wearable body sensors for patient self-assessment and monitoring. Arch Public Health 2014;72(1):28 [FREE Full text] [doi: 10.1186/2049-3258-72-28] [Medline: 25232478]
17. Rubeis G. The disruptive power of Artificial Intelligence. Ethical aspects of gerontechnology in elderly care. Arch Gerontol Geriatr 2020 Jul 15;91:104186. [doi: 10.1016/j.archger.2020.104186] [Medline: 32688106]
18. Ienca M, Wangmo T, Jotterand F, Kressig RW, Elger B. Ethical design of intelligent assistive technologies for dementia: a descriptive review. Sci Eng Ethics 2018 Aug;24(4):1035-1055. [doi: 10.1007/s11948-017-9976-1] [Medline: 28940133]
19. Ganyo M, Dunn M, Hope T. Ethical issues in the use of fall detectors. Ageing Soc 2011 Feb 07;31(8):1350-1367. [doi: 10.1017/s0144686x10001443]
20. Brownsell S, Hawley MS. Automatic fall detectors and the fear of falling. J Telemed Telecare 2004;10(5):262-266. [doi: 10.1258/1357633042026251] [Medline: 15494083]
21. DeCew J. Privacy. The Stanford Encyclopedia of Philosophy (Spring 2018 Edition). 2018. URL: https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=privacy [accessed 2021-02-12]
22. Mittelstadt B, Fairweather N, Shaw M, McBride N. Ethical issues of personal health monitoring: a literature review. The Research Center on Values in Emerging Science and Technology. URL: https://rcvest.southernct.edu/ethical-issues-of-personal-health-monitoring-a-literature-review/ [accessed 2021-02-01]
23. Tochen D. Wearable devices adoption poised to hit the mainstream. YouGovAmerica. URL: https://today.yougov.com/topics/technology/articles-reports/2016/08/15/wearable-devices-adoption-poised-growth-despite-la [accessed 2021-01-20]
24. Fadrique LX, Rahman D, Vaillancourt H, Boissonneault P, Donovska T, Morita PP. Overview of policies, guidelines, and standards for active assisted living data exchange: thematic analysis. JMIR Mhealth Uhealth 2020 Jun 22;8(6):e15923 [FREE Full text] [doi: 10.2196/15923] [Medline: 32568090]
25. Hecht A, Ma S, Porszasz J, Casaburi R, COPD Clinical Research Network. Methodology for using long-term accelerometry monitoring to describe daily activity patterns in COPD. COPD 2009 Apr;6(2):121-129 [FREE Full text] [doi: 10.1080/15412550902755044] [Medline: 19378225]
26. Van Hamme T, Garofalo G, Argones Rúa E, Preuveneers D, Joosen W. A systematic comparison of age and gender prediction on IMU sensor-based gait traces. Sensors (Basel) 2019 Jul 04;19(13):2945 [FREE Full text] [doi: 10.3390/s19132945] [Medline: 31277389]
27. Zwijsen SA, Niemeijer AR, Hertogh CM. Ethics of using assistive technology in the care for community-dwelling elderly people: an overview of the literature. Aging Ment Health 2011 May;15(4):419-427. [doi: 10.1080/13607863.2010.543662] [Medline: 21500008]
28. Marschollek M, Becker M, Bauer JM, Bente P, Dasenbrock L, Elbers K, et al. Multimodal activity monitoring for home rehabilitation of geriatric fracture patients--feasibility and acceptance of sensor systems in the GAL-NATARS study. Inform Health Soc Care 2014 Aug 22;39(3-4):262-271. [doi: 10.3109/17538157.2014.931852] [Medline: 25148561]
29. DeGrazia D, Millum J. A Theory of Bioethics. Cambridge, United Kingdom: Cambridge University Press; 2021.
30. Rubeis G, Schochow M, Steger F. Patient autonomy and quality of care in telehealthcare. Sci Eng Ethics 2018 Feb;24(1):93-107. [doi: 10.1007/s11948-017-9885-3] [Medline: 28281148]
31. Chiauzzi E, Rodarte C, DasMahapatra P. Patient-centered activity monitoring in the self-management of chronic health conditions. BMC Med 2015 Apr 09;13:77 [FREE Full text] [doi: 10.1186/s12916-015-0319-2] [Medline: 25889598]
32. Neuhauser L, Kreps GL, Morrison K, Athanasoulis M, Kirienko N, Van Brunt D. Using design science and artificial intelligence to improve health communication: chronologyMD case example. Patient Educ Couns 2013 Aug;92(2):211-217. [doi: 10.1016/j.pec.2013.04.006] [Medline: 23726219]
33. Essén A. The two facets of electronic care surveillance: an exploration of the views of older people who live with monitoring devices. Soc Sci Med 2008 Jul;67(1):128-136. [doi: 10.1016/j.socscimed.2008.03.005] [Medline: 18396367]
34. Kang HG, Mahoney DF, Hoenig H, Hirth VA, Bonato P, Hajjar I, Center for Integration of MedicineInnovative Technology Working Group on Advanced Approaches to Physiologic Monitoring for the Aged. In situ monitoring of health in older adults: technologies and issues. J Am Geriatr Soc 2010 Aug;58(8):1579-1586. [doi: 10.1111/j.1532-5415.2010.02959.x] [Medline: 20646105]
35. Wangmo T, Lipps M, Kressig RW, Ienca M. Ethical concerns with the use of intelligent assistive technology: findings from a qualitative study with professional stakeholders. BMC Med Ethics 2019 Dec 19;20(1):98 [FREE Full text] [doi: 10.1186/s12910-019-0437-z] [Medline: 31856798]
36. Sävenstedt S, Sandman PO, Zingmark K. The duality in using information and communication technology in elder care. J Adv Nurs 2006 Oct;56(1):17-25. [doi: 10.1111/j.1365-2648.2006.03975.x] [Medline: 16972914]

37. Baker R. Gait analysis methods in rehabilitation. J Neuroeng Rehabil 2006 Mar 02;3:4 [FREE Full text] [doi: 10.1186/1743-0003-3-4] [Medline: 16512912]
38. Chen D, Cai Y, Qian X, Ansari R, Xu W, Chu K, et al. Bring gait lab to everyday life: gait analysis in terms of activities of daily living. IEEE Internet Things J 2020 Feb;7(2):1298-1312. [doi: 10.1109/jiot.2019.2954387]
39. Becker C, Schwickert L, Mellone S, Bagalà F, Chiari L, Helbostad JL, FARSEEING Consortium, FARSEEING Meta Database Consensus Group. Proposal for a multiphase fall model based on real-world fall recordings with body-fixed sensors. Z Gerontol Geriatr 2012 Dec;45(8):707-715. [doi: 10.1007/s00391-012-0403-6] [Medline: 23184296]
40. Pol M, van Nes F, van Hartingsveldt M, Buurman B, de Rooij S, Kröse B. Older people's perspectives regarding the use of sensor monitoring in their home. Gerontologist 2016 Jun;56(3):485-493. [doi: 10.1093/geront/gnu104] [Medline: 25384761]
41. Demiris G, Hensel BK, Skubic M, Rantz M. Senior residents' perceived need of and preferences for "smart home" sensor technologies. Int J Technol Assess Health Care 2008;24(1):120-124. [doi: 10.1017/S0266462307080154] [Medline: 18218177]
42. Coventry L, Branley D. Cybersecurity in healthcare: a narrative review of trends, threats and ways forward. Maturitas 2018 Jul;113:48-52. [doi: 10.1016/j.maturitas.2018.04.008] [Medline: 29903648]
43. Wild K, Sharma N, Mattek N, Karlawish J, Riley T, Kaye J. Application of in-home monitoring data to transition decisions in continuing care retirement communities: usability study. J Med Internet Res 2021 Jan 13;23(1):e18806 [FREE Full text] [doi: 10.2196/18806] [Medline: 33439144]
44. Gabriels K, Moerenhout T. Exploring entertainment medicine and professionalization of self-care: interview study among doctors on the potential effects of digital self-tracking. J Med Internet Res 2018 Jan 12;20(1):e10 [FREE Full text] [doi: 10.2196/jmir.8040] [Medline: 29330140]
45. Steele R, Lo A, Secombe C, Wong YK. Elderly persons' perception and acceptance of using wireless sensor networks to assist healthcare. Int J Med Inform 2009 Dec;78(12):788-801. [doi: 10.1016/j.ijmedinf.2009.08.001] [Medline: 19717335]
46. Coughlin J, D'Ambrosio L, Reimer B, Pratt M. Proceedings of the 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2007 Presented at: 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Aug 22-26, 2007; Lyon, France. [doi: 10.1109/iembs.2007.4352665]
47. Bakken S, Reame N. The promise and potential perils of big data for advancing symptom management research in populations at risk for health disparities. Annu Rev Nurs Res 2016;34:247-260. [doi: 10.1891/0739-6686.34.247] [Medline: 26673385]
48. Courtney KL, Demiris G, Rantz M, Skubic M. Needing smart home technologies: the perspectives of older adults in continuing care retirement communities. Inform Prim Care 2008;16(3):195-201 [FREE Full text] [doi: 10.14236/jhi.v16i3.694] [Medline: 19094406]
49. Habibipour A, Padyab A, Ståhlbröst A. Social, ethical and ecological issues in wearable technologies. In: Proceedings of the Association for Information Systems. 2019 Presented at: AMCIS 2019 Proceedings, Association for Information Systems; Aug 15-17, 2019; Cancún, Mexico.
50. Yang K, Isaia B, Brown LJ, Beeby S. E-textiles for healthy ageing. Sensors (Basel) 2019 Oct 15;19(20):4463 [FREE Full text] [doi: 10.3390/s19204463] [Medline: 31618875]
51. Der Einfluss Künstlicher Intelligenz auf Freiheit, Gerechtigkeit und Solidarität. D64 Zentrum für Digitalen Fortschritt. URL: https://d-64.org/wp-content/uploads/2020/03/D64-Grundwerte-KI.pdf [accessed 2020-11-20]
52. Allen A. Privacy and medicine. Stanford Encyclopedia of Philosophy. 2009. URL: https://plato.stanford.edu/entries/privacy-medicine/ [accessed 2021-02-12]
53. Townsend D, Knoefel F, Goubran R. Privacy versus autonomy: a tradeoff model for smart home monitoring technologies. In: Proceedings of the 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2011 Presented at: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Aug 30- Sep 3, 2011; Boston, MA, USA. [doi: 10.1109/iembs.2011.6091176]
54. Azodo I, Williams R, Sheikh A, Cresswell K. Opportunities and challenges surrounding the use of data from wearable sensor devices in health care: qualitative interview study. J Med Internet Res 2020 Oct 22;22(10):e19542 [FREE Full text] [doi: 10.2196/19542] [Medline: 33090107]
55. Grates MG, Heming A, Vukoman M, Schabsky P, Sorgalla J. New perspectives on user participation in technology design processes: an interdisciplinary approach. Gerontologist 2019 Jan 09;59(1):45-57. [doi: 10.1093/geront/gny112] [Medline: 30260407]
56. Fang ML, Canham SL, Battersby L, Sixsmith J, Wada M, Sixsmith A. Exploring privilege in the digital divide: implications for theory, policy, and practice. Gerontologist 2019 Jan 09;59(1):e1-15. [doi: 10.1093/geront/gny037] [Medline: 29750241]
57. Chesser A, Burke A, Reyes J, Rohrberg T. Navigating the digital divide: a systematic review of eHealth literacy in underserved populations in the United States. Inform Health Soc Care 2016;41(1):1-19. [doi: 10.3109/17538157.2014.948171] [Medline: 25710808]
58. Oh SS, Kim K, Kim M, Oh J, Chu SH, Choi J. Measurement of digital literacy among older adults: systematic review. J Med Internet Res 2021 Feb 3;23(2):e26145. [doi: 10.2196/26145]

59. Peel NM, Paul SK, Cameron ID, Crotty M, Kurrle SE, Gray LC. Promoting activity in geriatric rehabilitation: a randomized controlled trial of accelerometry. PLoS One 2016 Aug 26;11(8):e0160906 [FREE Full text] [doi: 10.1371/journal.pone.0160906] [Medline: 27564857]
60. Landau R, Auslander GK, Werner S, Shoval N, Heinik J. Families' and professional caregivers' views of using advanced technology to track people with dementia. Qual Health Res 2010 Mar;20(3):409-419. [doi: 10.1177/1049732309359171] [Medline: 20133506]
61. Ethics of wearable medical and fitness devices. Technical University Munich. URL: https://benediktsvogler.com/downloads/Ethics%20of%20wearable%20medical%20and%20fitness%20devices.pdf [accessed 2020-11-21]
62. Lim SE, Dodds R, Bacon D, Sayer AA, Roberts HC. Physical activity among hospitalised older people: insights from upper and lower limb accelerometry. Aging Clin Exp Res 2018 Nov;30(11):1363-1369 [FREE Full text] [doi: 10.1007/s40520-018-0930-0] [Medline: 29542070]
63. von Storch K, Schlomann A, Rietz C, Polidori MC, Woopen C. [Wearables to support self-management of older adults with chronic diseases : a qualitative study from the perspectives of patients and physicians]. Z Gerontol Geriatr 2018 Nov;51(7):791-798. [doi: 10.1007/s00391-017-1323-2] [Medline: 28986628]
64. Van Velsen L, Wildevuur S, Flierman I, Van Schooten B, Tabak M, Hermens H. Trust in telemedicine portals for rehabilitation care: an exploratory focus group study with patients and healthcare professionals. BMC Med Inform Decis Mak 2016 Jan 27;16:11 [FREE Full text] [doi: 10.1186/s12911-016-0250-2] [Medline: 26818611]

Edited by L Buis; submitted 14.08.21; peer-reviewed by P Bhadra, J Salisbury; comments to author 19.10.21; revised version received 27.11.21; accepted 22.02.22; published 23.06.22.

Please cite as:
Predel C, Timmermann C, Ursin F, Orzechowski M, Ropinski T, Steger F
Conflicting Aims and Values in the Application of Smart Sensors in Geriatric Rehabilitation: Ethical Analysis
JMIR Mhealth Uhealth 2022;10(6):e32910
URL: https://mhealth.jmir.org/2022/6/e32910
doi:10.2196/32910
PMID:35737429

Original Paper

Fitbit Use and Activity Levels From Intervention to 2 Years After: Secondary Analysis of a Randomized Controlled Trial

Sheri J Hartman[1,2], PhD; Ruohui Chen[1,2], MSc; Rowena M Tam[1], PT, DPT; Hari K Narayan[2,3], MD, MSCE; Loki Natarajan[1,2], PhD; Lin Liu[1,2], PhD

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

[2]Moores Cancer Center, University of California, San Diego, San Diego, CA, United States

[3]Department of Pediatrics, University of California, San Diego, San Diego, CA, United States

Corresponding Author:
Sheri J Hartman, PhD
Herbert Wertheim School of Public Health and Human Longevity Science
University of California, San Diego
3855 Health Sciences Drive, #0901
La Jolla, CA, 92037
United States
Phone: 1 8585349235
Email: sjhartman@ucsd.edu

Abstract

Background: There has been a rapid increase in the use of commercially available activity trackers, such as Fitbit, in physical activity intervention research. However, little is known about the long-term sustained use of trackers and behavior change after short-term interventions.

Objective: This study aims to use minute-level data collected from a Fitbit tracker for up to 2 years after the end of a randomized controlled trial to examine patterns of Fitbit use and activity over time.

Methods: Participants in this secondary data analysis were 75 female breast cancer survivors who had been enrolled in a 12-week physical activity randomized controlled trial. Participants randomized to the exercise intervention (full intervention arm) received a Fitbit One, which was worn daily throughout the 12-week intervention, and then were followed for 2 years after the intervention. Participants randomized to the waitlist arm, after completing the randomized controlled trial, received a Fitbit One and a minimal version of the exercise intervention (light intervention arm), and then were followed for 2 years after the intervention. Average and daily adherence and MVPA were compared between the 2 groups in the interventional and postinterventional periods using both linear and generalized additive mixed effects models.

Results: Adherence to wearing the Fitbit during the 12-week intervention period was significantly higher in the full intervention arm than in the light intervention arm (85% vs 60%; $P<.001$). Average adherence was significantly lower for both study arms during the follow-up period than in the intervention period; however, there were statistically different patterns of adherence during the follow-up period, with the light intervention arm having steeper declines than the full intervention arm over time ($P<.001$). Similar to the adherence results, mean minutes of Fitbit-measured MVPA was higher for the full intervention arm than for the light intervention arm during the 12-week intervention period (mean MVPA 27.89 minutes/day, SD 16.38 minutes/day vs 18.35 minutes/day, SD 12.64 minutes/day; $P<.001$). During the follow-up period, average MVPA was significantly lower than the 12-week intervention period for both the full intervention arm (21.74 minutes/day, SD 24.65 minutes/day; $P=.002$) and the light intervention arm (15.03 minutes/day, SD 13.27 minutes/day; $P=.004$). Although the mean MVPA in each arm was similar across the follow-up period ($P=.33$), the pattern of daily MVPA was significantly different between the 2 groups ($P<.001$).

Conclusions: While adherence to wearing activity trackers and maintaining physical activities declined after completion of a 12-week exercise intervention, a more active interventional strategy resulted in greater wear time and activity levels during the intervention and more stable patterns of adherence and activity in the long term. An improved understanding of long-term maintenance patterns may inform improved exercise interventions that result in sustained increases in physical activity.

Trial Registration: ClinicalTrials.gov NCT02332876; https://clinicaltrials.gov/ct2/show/NCT02332876

(JMIR Mhealth Uhealth 2022;10(6):e37086) doi:10.2196/37086

KEYWORDS

physical activity; fitness; exercise; Fitbit; wearable; health technology; mHealth; digital health; activity tracker; maintenance; adherence; tracker; use pattern; activity level; behavior change; cancer; breast; survivor; long-term use; sustained use

Introduction

There are currently 3.9 million breast cancer survivors in the United States; most of whom do not engage in sufficient physical activity to meet current recommendations [1]. Greater physical activity in breast cancer survivors is associated with better quality of life, lower risk of all-cause and breast cancer–specific mortality, and lower risk of recurrent breast cancer [2-6], but 34% of cancer survivors report no physical activity in their leisure time [7,8]. An abundance of evidence demonstrates that interventions to increase physical activity in breast cancer survivors can be effective in the short term [4,5]. However, there are few studies examining maintenance of longer-term physical activity behavior beyond the intervention period [9] and those that do suggest that physical activity declines over time [10-12]. An improved understanding of maintenance behaviors is needed to optimize interventions to sustain increases in physical activity over the long term. Wearable trackers, such as Fitbit, capture physical activity behaviors and provide self-monitoring feedback, thereby offering both greater insight into maintenance behaviors and a potential method to facilitate sustained improvements in long-term maintenance.

Self-monitoring is one of the key skills to promote behavior changes [13], and may have a role in promoting sustained increases in physical activity in breast cancer survivors. The behavior change techniques framework proposed by Michie and colleagues [13,14] suggests that self-monitoring is the skill most strongly associated with intervention success when combined with at least one other self-regulatory technique from Control Theory (eg, receiving feedback on performance and reviewing progress toward goals) [15,16]. According to Control Theory, feedback loops provide awareness of discrepancies between performance and goals that can encourage behavior change [15]. Wearable trackers facilitate self-monitoring and feedback loops by passively collecting and providing information and feedback on progress toward individual goals.

Initial studies on Fitbit adoption have demonstrated that they are effective in increasing physical activity levels when coupled with other interventions [17-21], but the novelty of wearing the tracker wears off over time [22]. Additionally, prior studies have either been short term or had continued contacts with the participants in their maintenance phase [22-24]. Studies that have only utilized Fitbit as a means of behavior change show no significant changes in physical activity [25-27]. This decline in novelty, short interventional period, and variable additional support throughout the intervention may negatively affect use of the wearable technology when external accountability from the research study is removed [28-31].

This analysis explored adherence to wearing the Fitbit and physical activity 2 years after the end of a 3-month randomized controlled trial comparing a physical activity intervention (full intervention arm) with a waitlist control that received a “light” intervention (light intervention arm) after completing the 12-week assessments [24]. The aims of this study were to (1) examine patterns of adherence to wearing the Fitbit between the full intervention arm and the light intervention arm during their respective 3-month intervention periods and up to 2 years’ follow-up; (2) examine patterns of Fitbit-measured physical activity between the full intervention arm and the light intervention arm during their respective 3-month intervention periods and up to 2 years’ follow-up.

Methods

Participants and Design

Participants in this secondary data analysis were originally randomized to a 12-week physical activity intervention group or a waitlist control group. After completing final measures for the randomized trial at week 12, participants were invited to enroll in a maintenance study where their Fitbit data would be collected for the next 2 years and they would complete online questionnaires every 6 months over the next 2 years (4 times total). Participants were asked to provide written informed consent for participating in the maintenance study. Data from the original randomized trial and the 2-year follow-up were collected from February 2015 to July 2018. The intervention trial was registered with Clinicaltrials.gov (NCT02332876).

Eligible participants were female breast cancer survivors, aged 21-85 years, who were diagnosed less than 5 years prior to study enrollment, had completed chemotherapy or radiation treatment, were sedentary (defined as self-reporting <60 minutes of moderate-to-vigorous physical activity [MVPA] in 10-minute bouts per week), and had access to the internet and a Fitbit-compatible computer, tablet, or phone. Exclusion criteria included any medical condition that could make it potentially unsafe to be in an unsupervised physical activity intervention (determined by the Physical Activity Readiness Questionnaire [32]), other primary or recurrent invasive cancer within the last 10 years, and inability to commit to a 12-week intervention. All participants who returned for the 12-week assessment were eligible to enroll in the maintenance study.

A detailed description of the original trial’s protocol was previously published [33]. Briefly, potential participants were telephone screened, with interested and eligible women scheduled for an in-person visit to provide signed informed consent and complete baseline measures. Participants returned about 1 week later for their second visit where they were randomly assigned to 1 of 2 groups, an exercise intervention or waitlist control, in a 1:1 ratio. After randomization, participants in both groups reviewed the expectations and requirements of their group assignment with study staff.

Physical Activity Intervention (Full Intervention Arm)

Participants randomized to the full intervention arm had a 30- to 45-minute in-person meeting where they went on a 10-minute walk at moderate intensity and set personalized physical activity

goals with a researcher trained in motivational interviewing aimed at gradually working up to 150 minutes/week of MVPA. Participants were given a Fitbit One (Fitbit, Inc./Google) to self-monitor their physical activity, set up the Fitbit with their coach, taught how to use it, and taken on a 10-minute walk. Participants were also informed that their health coach would be reviewing their Fitbit activity data weekly and that they would receive feedback on the Fitbit data during the scheduled phone calls and between calls as needed. Participants received 2 scheduled phone calls (2- and 6-week time points) and emails every 3 days throughout the 12-week intervention. The intervention was delivered by a clinical psychologist with extensive training and experience in promoting behavior change (SJH) and by a staff member who was trained by SJH. For further details on the intervention, see Hartman et al [33]. No additional intervention content and support were received during the 2-year follow-up period.

Waitlist Plus Light Physical Activity Intervention (Light Intervention Arm)

After completion of measures at the final visit, participants in the light intervention arm were provided with a "light" version of the exercise intervention. In a 15-20-minute in-person meeting, participants in the light intervention arm worked with a measurement research assistant to set personalized physical activity goals. The research assistant had received training on goal setting from SJH, with a brief introduction to using motivational interviewing. Participants received the Fitbit One with instructions on how to use it to support self-monitoring. Different from the full intervention arm, participants did not set up the Fitbit with their health coach, they were not told that their health coach could see their data nor that they would receive any feedback on their Fitbit data. Participants were also not taken on the 10-minute walk to demonstrate moderate intensity. Participants were offered the same 2 phone calls (2 and 6 weeks later), but these calls were framed as optional. Participants received the same automated emails every 3 days for the next 12 weeks that the full intervention arm received.

Two-Year Maintenance Study Assessments

At the completion of their respective intervention, participants in both arms were asked to sync and charge their Fitbit at least once per week. When participants had not synced their Fitbit for 2 weeks, study staff would contact them to ask them to sync and provide any tech support if there were challenges syncing. Participants also received 4 online questionnaires to complete every 6 months across the 2-year follow-up.

Measures

The Fitbit One, a commercially available accelerometer-based activity tracker, was used to examine patterns of physical activity throughout the 12-week intervention. Fitbit uses a proprietary algorithm to classify each minute as being in sedentary, light, moderate, or vigorous activity, and provides metabolic equivalents (METs) for each minute. Data were wirelessly uploaded to the user's Fitbit account online and then downloaded by the research team through a database called Fitabase (Small Steps Lab), which allows for collecting data at the minute level. Daily adherence to wearing the Fitbit tracker was defined as wearing the tracker for over 10 hours in a day or logging at least some activity (over 1 minute of MVPA). This definition for a valid Fitbit wear day was used because participants were not instructed to wear the Fitbit all day; rather they were instructed to use the Fitbit to track activity. Thus, wearing the tracker specifically to log MVPA was deemed to be valid wear based on these instructions. Fitbit wear time was determined by processing of minute-level Fitbit data using the R function *accel.weartime* within the "accelerometry" package [34]. Nonwear was classified using both steps and METs. Consistent with standard protocols for ActiGraph accelerometry wear time [35], greater than 90 consecutive minutes of 0 steps/METs with 2-minute tolerance (ie, for 2 minutes with nonzero counts during nonwear intervals) was deemed nonwear.

Both groups wore the ActiGraph for 7 days prior to receiving the Fitbit and starting the full or light intervention. The ActiGraph GT3X+, a well-validated research-grade accelerometer [36], provided frequency, duration, and intensity of physical activity. Using standard guidelines, sufficient ActiGraph wear time was classified as over 10 hours of wear a day for at least 5 days or over 50 hours across 4 days and screened for in the ActiLife software using guidelines outlined by Choi et al [35]. All complete and valid data were processed in the ActiLife software using the low-frequency extension and aggregated to 60-second epochs so that published physical activity cut points could be applied [37]. MVPA was defined as 1952 or more counts per minute (3.00-7.00 METs). The full intervention arm wore the Fitbit and ActiGraph concurrently for 7 days to assess validity of Fitbit-measured MVPA. Fitbit-measured MVPA was highly correlated with ActiGraph-measured MVPA collected on overlapping days (r=0.81: ActiGraph MVPA/day mean 29.9 minutes, SD 25.90 minutes; Fitbit MVPA/day mean 25.8 minutes, SD 28.76 minutes), as we have previously reported [24].

On the questionnaire administered at 6, 12, 18, and 24 months, participants were asked if they were still wearing their Fitbit. If they reported they were not wearing it, they were asked the reason they stopped wearing the Fitbit.

Statistical Analysis

Participants who did not consent for 2-year maintenance study were excluded from the analysis. Group differences in baseline characteristics between those who consented to the 2-year study and those who did not were assessed using 2-sample independent *t* test (unpaired) and chi-square test. Baseline characteristics were summarized between the full intervention arm and the light intervention arm.

Adherence to Wearing the Fitbit and Daily MVPA During the 12-Week Intervention Period and 2-Year Follow-Up

The mean weekly rolling average adherence to wearing the Fitbit and mean MVPA were calculated by averaging the outcomes over the first 12-week period and over the 2-year follow-up period separately for each individual. Descriptive statistics and boxplots were used to summarize the adherence to wearing the Fitbit and MVPA at 12-week and 2-year follow-up as well as the change from 12 weeks to 2 years.

Comparison of Adherence to Wearing the Fitbit and MVPA Between the Full Intervention Arm and the Light Intervention Arm

For comparing mean outcomes and mean change in outcomes between the 2 intervention groups, we used the following linear mixed effects model:

$$E(Y) = \beta_0 + \beta_1 \times Arm + \beta_2 \times Period + \beta_3 \times Arm \times Period + b_0 + b_1 \times Period$$

where Arm and Period are binary variables for the study arm (full or light intervention) and study period (12-week or 2-year follow-up), respectively; random intercept b_0 and random slope b_1 are included to account for correlation among repeated measures within each individual. The coefficient β_1 indicates the mean outcome difference between the full intervention arm and the light intervention arm at the first 12 weeks; $\beta_1 + \beta_3$ indicates the mean outcome difference between the 2 arms at 2-year follow-up; β_2 indicates the mean outcome change from the first 12-week and 2-year follow-up for the light intervention arm; $\beta_2 + \beta_3$ indicates the mean outcome change between the first 12-week and 2-year follow-up for the full intervention arm; β_3 indicates the difference in mean outcome change from the first 12-week and 2-year follow-up between the full intervention arm and the light intervention arm. The *P* value for testing the significance was calculated based on the estimated coefficient and estimated covariance from the linear mixed effects model.

To compare the trend of adherence and MVPA between the full intervention arm and the light intervention arm, we used the generalized additive mixed effects model (GAMM):

$$g(y) = \beta_0 + \beta_1 \times Arm + s(Time) + s(Time) \times Arm$$

where Time is a continuous variable for the study day (day 1, day 2, …); s(Time) is the smooth term for "Time"; and s(Time) × Arm is the interaction term between "Time" and "Arm." Models with and without prespecified knots were assessed.

We used the minimized generalized cross-validation score for smoothness selection. To select the best fitted model, in terms of the interaction term between time and group and knots specification in the GAMM, we conducted model comparisons using analysis of variance and model's Akaike information criteria. For the goodness of fit of the chosen models, we examined the model's deviance and the adjusted R^2. Graph of the best fit was used to display the trends of adherence and MVPA over the study period.

Comparison of MVPA Between Preintervention and Postintervention

We also used paired *t* test to compare the MVPA during the preintervention period (measured by ActiGraph) with MVPA during the 12-week intervention, and MVPA during the 2-year follow-up period for both the full intervention arm and the light intervention arm. We also compared the preintervention MVPA between the 2 study groups using the 2-sample independent *t* test (unpaired).

Ethics Approval

All procedures were approved by the University of California San Diego Human Subjects Protection Program (IRB#140694).

Results

Participant Characteristics

Of the 911 women who were screened for eligibility, 97 were eligible and scheduled for a visit, and 87 participants were randomized. Most common reasons for being ineligible were being too active (n=225), unable/unwilling to attend clinic visits (n=106), breast cancer surgery more than 5 years ago (n=81), and medical exclusion (n=36). Of the 87 randomized, 75 agreed to enroll in the 2-year maintenance study: 37/43 in the full intervention arm (86%) and 38/44 in the light intervention arm (86%). The current analyses comprise data from the 75 participants who enrolled in the maintenance study. There were no significant (*P*>.05) differences in demographic or clinical variables between participants who did and did not enroll in the maintenance study.

Participants in the 2-year follow-up study were 75 female breast cancer survivors who were predominantly diagnosed at stage 1. A little more than half had received chemotherapy and at the start of the original trial they were on average 2.6 years from the diagnosis. The average age of participants was 57 years (SD 10.4 years), with the majority being non-Hispanic, White, and having a college education or greater (Table 1). There were no significant (*P*>.05) differences between the 2 arms.

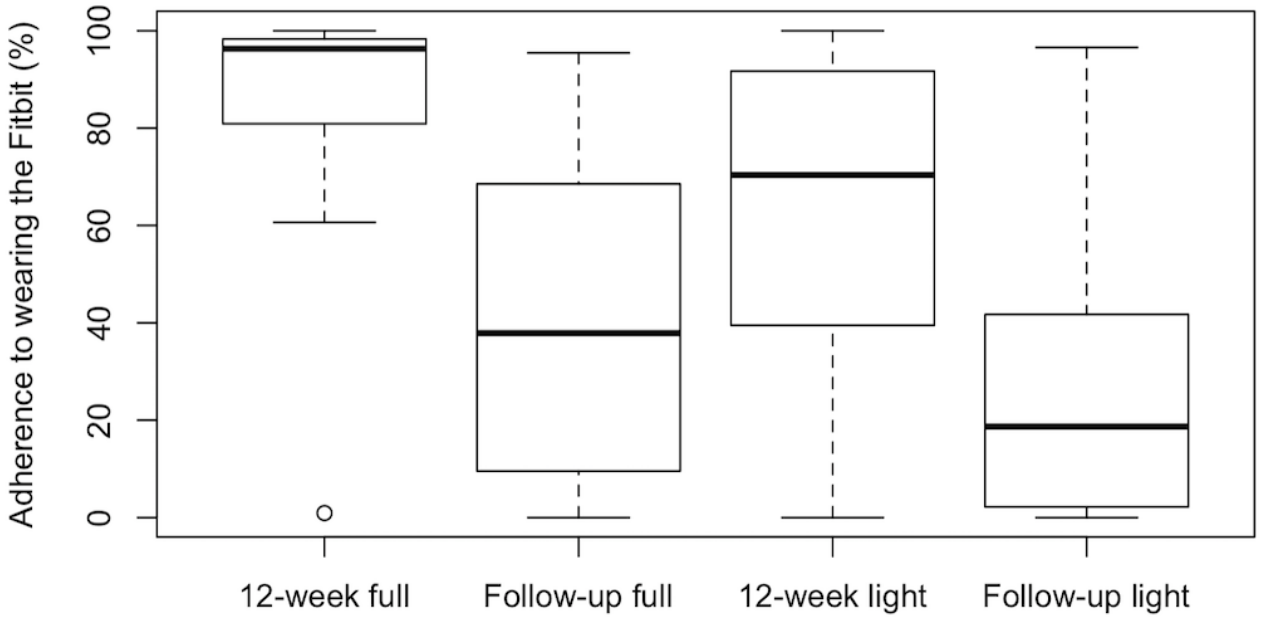
Table 1. Baseline characteristics by intervention group.

Characteristics	Full intervention (n=37)	Light intervention (n=38)	All (n=75)
Age (years), mean (SD)	58.2 (11.5)	56.2 (9.1)	57.2 (10.4)
Married status, n (%)	27 (72.9)	27 (71.1)	54 (72.0)
BMI (kg/m^2), mean (SD)	26.7 (6.4)	27.7 (6.4)	27.2 (6.4)
Education, n (%)			
Some college or less	11 (29.7)	9 (23.7)	20 (26.7)
College graduate	15 (40.6)	20 (52.6)	35 (46.7)
Master or higher	11 (29.7)	9 (23.7)	20 (26.7)
Ethnicity, n (%)			
Not Hispanic/Latino	30 (81.1)	33 (86.8)	63 (84.0)
Hispanic/Latino	7 (18.9)	5 (13.2)	12 (16.0)
Race, n (%)			
White	30 (80.1)	31 (81.6)	61 (81.3)
Non-White	7 (18.9)	7 (18.4)	14 (18.7)
Cancer stage, n (%)			
Stage 1	22 (59.4)	22 (57.9)	44 (58.7)
Stage 2	11 (29.7)	13 (34.2)	24 (32.0)
Stage 3	4 (10.8)	3 (7.9)	7 (9.3)
Received chemotherapy, n (%)	21 (56.7)	20 (52.6)	41 (54.7)
Time since surgery (months), mean (SD)	31.4 (17.0)	30.6 (16.0)	30.9 (16.4)

Patterns of Adherence to Wearing the Fitbit

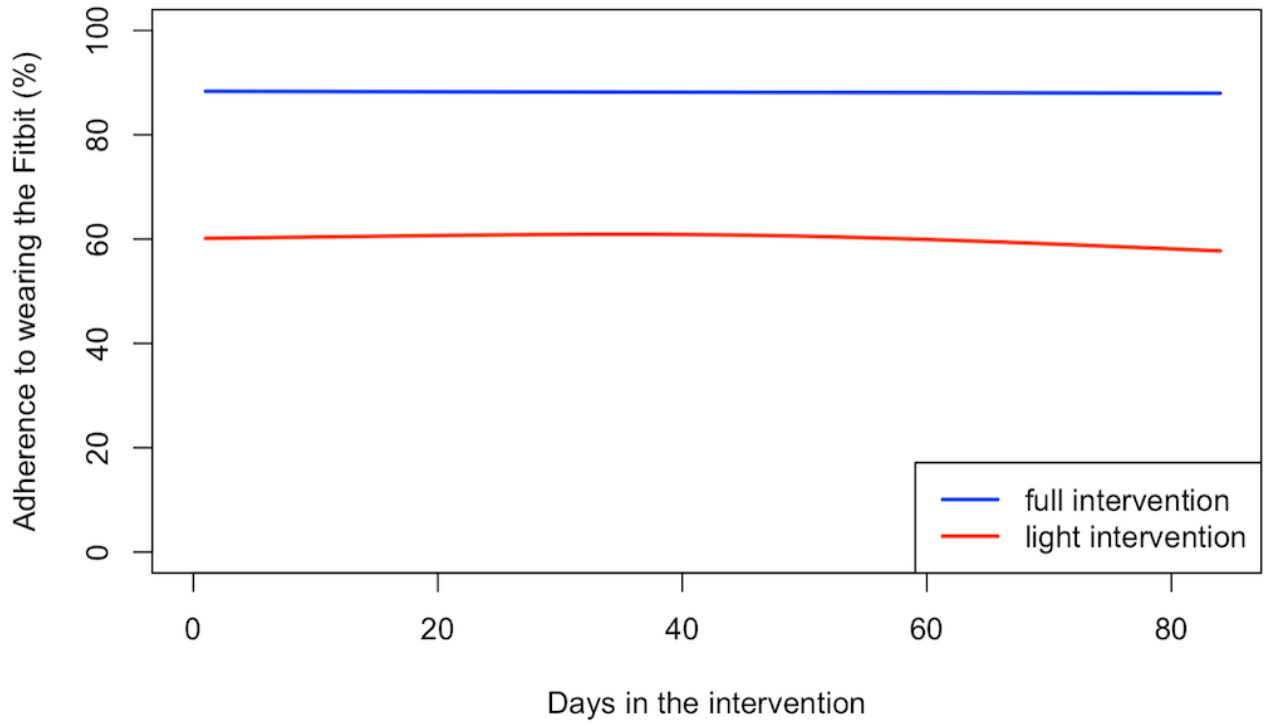
Average adherence to wearing the Fitbit was significantly higher for the full intervention arm during the 12-week intervention period compared with the light intervention arm during the 12-week intervention period—mean adherence 85% (SD 23%) for the full intervention arm versus 60% (SD 34%) for the light intervention arm (P<.001). In addition, average adherence from the postintervention to 2-year follow-up period significantly dropped from the 12-week intervention period for both the full intervention arm (45%, SD 33%; P<.001) and the light intervention arm (30%, SD 31%; P<.001). However, during the postintervention to 2-year period there were no significant differences in average adherence between the 2 groups (Figure 1)—mean adherence 40% (SD 35%) for the full intervention arm versus 30% (SD 32%) for the light intervention arm (P=.71).

Figure 1. Box-plots of median and interquartile range of adherence to wearing the Fitbit during the 12-week exercise intervention or "light" intervention period for each study group, and during the post-intervention to 2-year follow-up period for each study group.



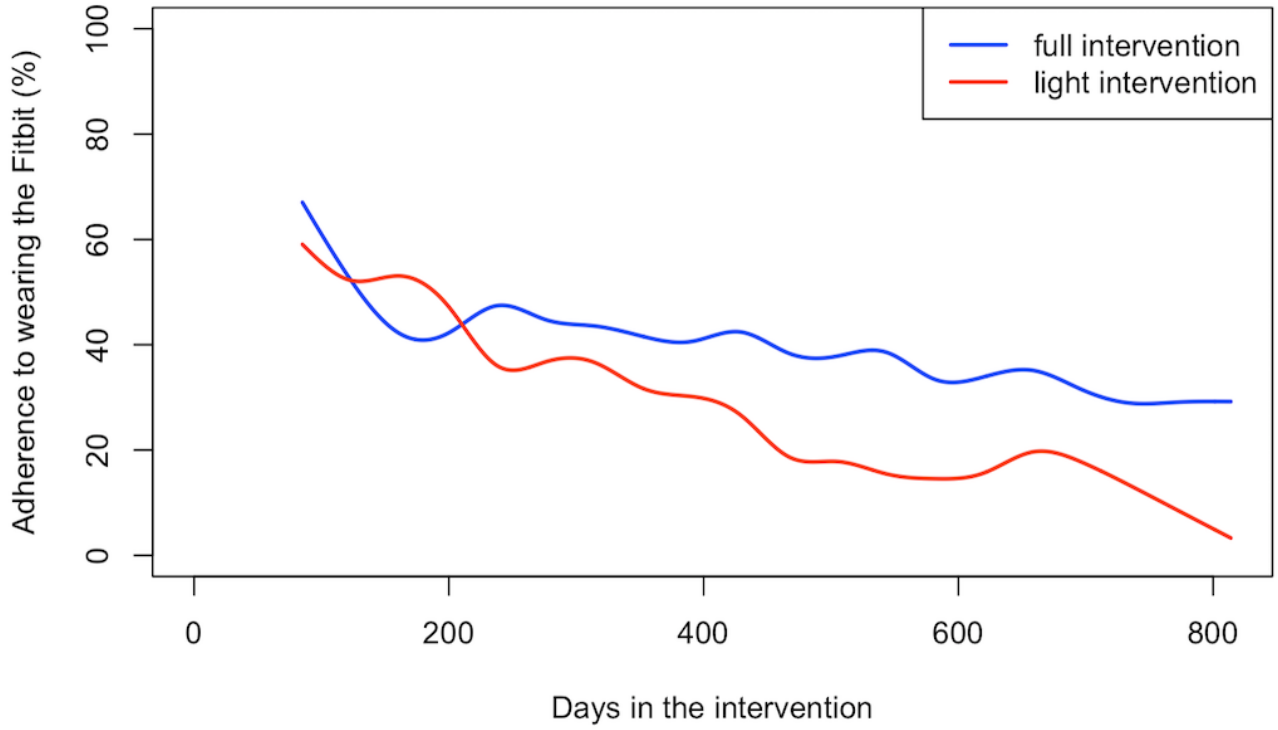
We then compared the temporal patterns of adherence between the 2 groups during the 12-week intervention period and the postintervention to 2-year period using the GAMM. While participants in the full intervention arm had significantly higher (P<.001) average adherence during the 12-week interventional period (Figure 2), there was no significant difference in the temporal pattern of adherence across the 12-week period (P=.24), with both groups having stable adherence over time.

Figure 2. Weekly rolling average adherence to wearing the Fitbit during the 12-week intervention period for the Full Intervention arm and the 12-week "light" intervention period for the Light Intervention arm, by group.



By contrast, in the postintervention to 2-year period (Figure 3), although the average adherence across the entire postintervention period was similar, the daily adherence over time was significantly different between the 2 groups (P<.001). While there were steep initial declines in both arms, adherence in the full intervention arm declined more gradually over the remainder of the study period in comparison with the light intervention arm.

Figure 3. Weekly rolling average adherence to wearing the Fitbit after completion of the 12-week intervention period, by group.



A total of 67 participants answered the self-report question of whether or not they were still wearing their Fitbit. Of these, 32 participants reported that they had stopped wearing their Fitbit: 13 (41%) stated that their Fitbit broke, 12 (38%) reported that they lost their Fitbit or their charger, 6 (19%) stated that they were no longer interested in wearing the Fitbit, and 3 (9%) indicated a health issue that stopped them from being active. Several participants replaced lost or broken Fitbits during the follow-up years and then subsequently had a lost or broken Fitbit a second time or lost interest in wearing it.

Patterns of Fitbit-Measured MVPA

Participants in the full intervention arm significantly increased average minutes of MVPA from preintervention to across the 12-week intervention period (13.95 minutes/week to 27.89 minutes/week; *P*<.001) and participants in the light intervention arm showed a trend for increased average minutes of MVPA from preintervention to across the 12 weeks (14.64 minutes/week to 18.35 minutes/week; *P*=.07; Table 2). Although both arms increased MVPA during the 12-week intervention period, the full intervention arm had significantly higher average minutes of MVPA than the light intervention arm (27.89 minutes/week versus 18.35 minutes/week, respectively; *P*<.001; Table 2). Similar to the adherence results, during the 2-year postintervention period the average MVPA significantly dropped from the 12-week intervention period for both the full intervention arm (21.74 minutes/week at the 2-year follow-up vs 27.89 minutes/week at the 12-week intervention; *P*=.002) and the light intervention arm (15.03 minutes/week at the 2-year follow-up vs 18.35 minutes/week at the 12-week intervention; *P*=.004), but there was no significant difference in average MVPA between the 2 groups (*P*=.33). Although average MVPA decreased during the 2-year follow-up in comparison to preintervention MVPA, there was a trend for greater average MVPA for participants in the full intervention arm (21.74 minutes/week at the 2-year follow-up vs 13.95 minutes/week preintervention; *P*=.08), but no difference from preintervention for the light intervention arm (15.03 minutes/week at the 2-year follow-up vs 14.64 minutes/week preintervention; *P*=.26).

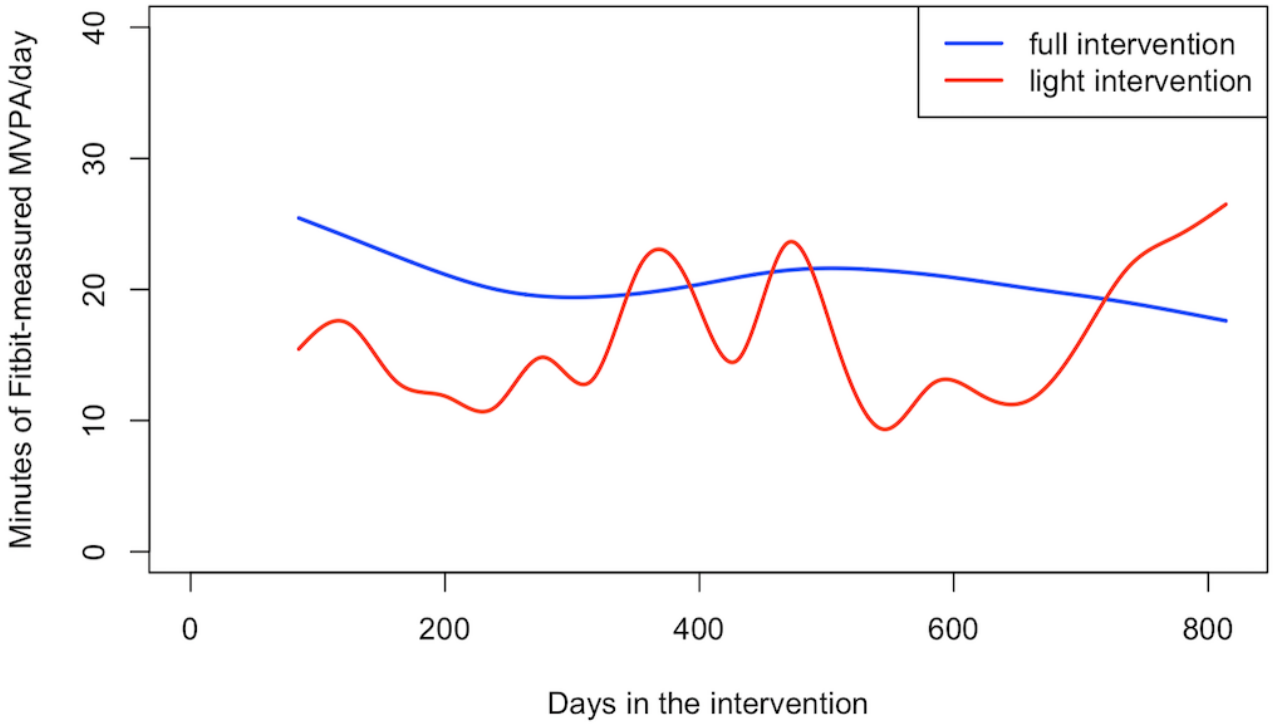
Table 2. Minutes per day of moderate-to-vigorous physical activity, by group (N=75).

Physical activity	Full intervention MVPA (minutes/week), mean (SD)	*P* value for comparison of preintervention with postintervention within the full intervention group	Light intervention MVPA (minutes/week), mean (SD)	*P* value for comparison of preintervention with postintervention within the light intervention group	*P* value for comparison between groups
Preintervention (ActiGraph)	13.95 (11.96)	N/A[a]	14.64 (13.46)	N/A	.83
12-week intervention period (Fitbit)	27.89 (16.38)	<.001	18.35 (12.64)	.07	<.001
2-year follow-up (Fitbit)	21.74 (24.65)	.08	15.03 (13.27)	.26	.33

[a]N/A: not applicable.

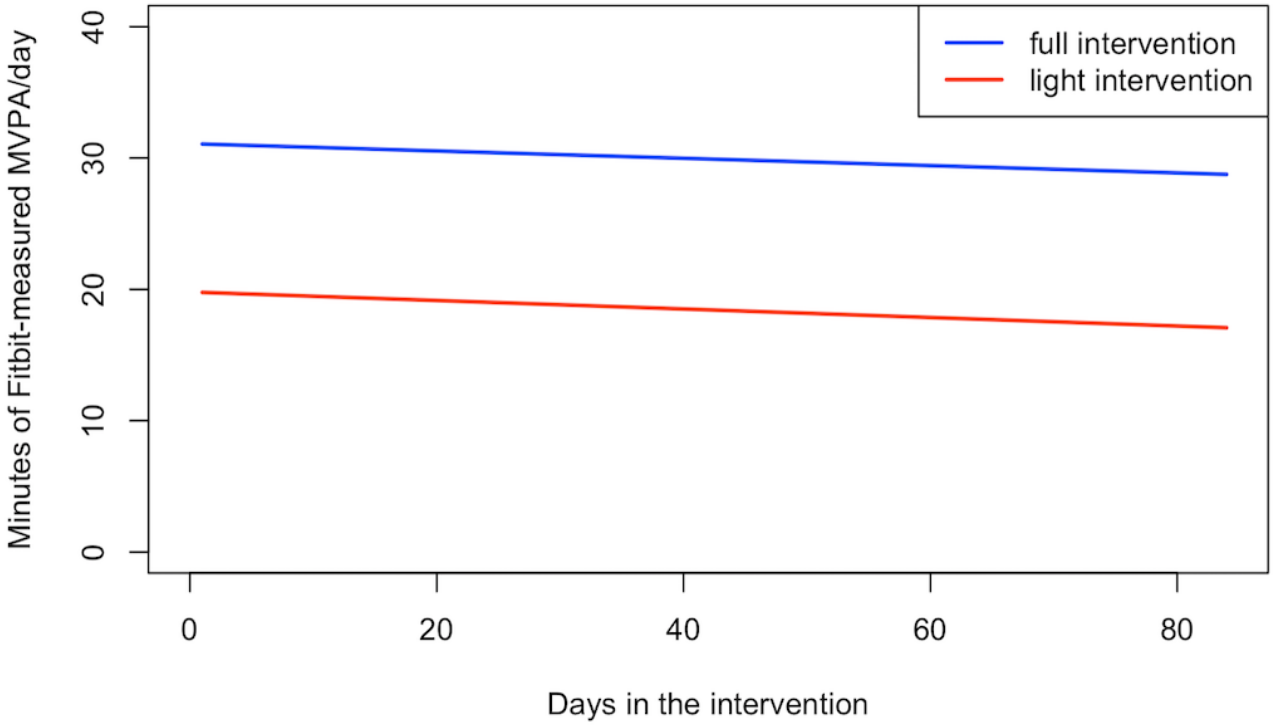
We then compared the temporal patterns of activity between the 2 groups during the 12-week intervention period and from the postintervention to 2-year period (Figure 4). Similar to the adherence results, while the full intervention arm had significantly higher (P=.002) daily MVPA than the light intervention arm during the intervention, there was no difference in the temporal pattern of daily MVPA across the 12-week period (P=.99), with both groups having relatively stable daily MVPA.

Figure 4. Daily Fitbit measured MVPA during the 12-week intervention period for the Full Intervention arm and the 12-week "light" intervention period for the Light Intervention arm, by group.



While the average MVPA for the entire postintervention period was similar between groups, the daily MVPA over time was significantly different between the 2 groups (Figure 5; P<.001). Among participants who continued to adhere to wearing the Fitbit, the full intervention arm had a relatively stable trend with a gradual decline in daily MVPA, while the light intervention arm had an irregular temporal pattern with fluctuations in MVPA over time. Of note, with the relatively low level of adherence that continued to decline over the follow-up period, the curvature trend of the daily MVPA in the light intervention arm was measured in a small number of individuals.

Figure 5. Daily Fitbit measured MVPA after completion of the 12-week intervention period, by group.



Discussion

Principal Findings

This study examined patterns of wearing an activity tracker and engaging in MVPA during and after completion of a 12-week randomized trial of a full exercise intervention in comparison with a light intervention, through 2 years of follow-up. Our study yielded several key findings. First, the exercise intervention, which entailed more comprehensive feedback and external accountability, resulted in greater adherence to wearing the Fitbit and minutes of MVPA than the light intervention. Second, both full and light interventional groups had significant reductions in adherence and physical activity during the long-term follow-up. Finally, while both groups had similar average adherence and MVPA during the postintervention to 2-year period, the full intervention group had a more stable temporal pattern of adherence and daily MVPA during this time than the light intervention group, in addition to a trend toward maintaining some gains in MVPA over preintervention levels. These results provide new insight into wearable technology and activity patterns during and after completion of an exercise intervention and suggest the potential importance of sustained self-monitoring and feedback interventions to maintain increased activity levels over time.

We found that the full exercise intervention resulted in greater daily MVPA in comparison with the light intervention during the 12-week interventional period. The primary added feature of the full exercise intervention was external accountability, where participants were aware that their activity would be checked by their health coach, discussed with them at planned phone calls, and would receive additional contacts in between calls based on their Fitbit data. Our results suggest that this accountability led to greater adherence to wearing the Fitbit and MVPA during the intervention period. Wearing the Fitbit did not decrease over time during the 12-week intervention period for either study arm. This may have been due to the external cues and reminders to wear and sync their Fitbit that participants received from the emails that came 2-3 times a week during that period, but stopped at 12 weeks, or may have been due to the novelty of using a Fitbit. However, wearing the Fitbit decreased after the intervention period with the most commonly reported reasons for discontinued use of the Fitbit were that it broke or was lost. Now that most Fitbits are wrist-worn, it may help to decrease loss of devices, but devices breaking is likely to be a continued issue that impacts continued wearing of devices. With the well-established benefits of self-monitoring for behavior change, identifying ways to increase long-term engagement with activity trackers is needed.

We also found that both the full and the light intervention group increased minutes of MVPA from the preintervention to the intervention period and maintained it during the 12-week intervention period. The initial increase in MVPA at the start rather than gradually increasing overtime may have been due to the intervention's goal-setting approach that utilized motivational interviewing, where participants were allowed to set any starting goal that they chose. With more studies having day-level physical activity data, future studies could examine different methods of setting goals and different patterns of exercise to see if they relate to long-term maintenance of activity. This study adds to the literature by demonstrating the importance of additional intervention components, particularly increasing external accountability, when using activity trackers to promote exercise, and the challenges with lost and broken trackers.

Our study is one of the first to use wearable activity trackers to assess long-term maintenance of behavior after completion of

a short-term intervention in breast cancer survivors. Although some prior studies have assessed long-term physical activity, they have either examined sustained long-term physical activity interventions or relied upon self-reported MVPA [9]. In our study, there were significant declines in wearing the Fitbit and activity levels after the end of the intervention. This is consistent with previous studies in cancer survivors that have generally found that activity levels reduce from the end of the interventional period [9]. Our results suggest that simply allowing participants to keep a wearable tracker is not sufficient to maintain activity levels in the long term. As the novelty of having the tracker wanes, additional measures, such as continued external accountability or coaching, may be beneficial.

Although there were significant declines in activity levels during the postintervention follow-up for both groups, it is interesting that the temporal patterns of both adherence and physical activity were more stable in those who received the full exercise intervention. Comparison of these results with previous trials is difficult as this study took advantage of having minute-level physical activity data for months, rather than having brief snapshots of MVPA from 7 days of accelerometer wear or self-report. By examining patterns of activity over time we see that even a short-term intensive exercise intervention may result in some lasting change in behavior patterns beyond the intervention period. Together, the results suggest that an intensive short-term physical activity intervention, coupled with a continued long-term maintenance intervention, may be necessary to sustain higher activity levels in the long term. Further study is needed to develop the optimal short- and long-term strategies to enhance activity tracker use to achieve sustained physical activity.

Limitations

This study provides unique insight into long-term activity levels after completion of an exercise intervention, but there are several limitations that should be noted. The sample size limited our ability to detect potentially smaller differences between groups, including average adherence and MVPA in the postintervention period. In addition, the progressive decline in adherence to wearing the Fitbit in long-term follow-up meant that there was a large amount of missing MVPA data. Without other measures of MVPA we are unable to know how much activity individuals were engaging in after their Fitbit broke, was lost, or if they were no longer interested in wearing, and our results are limited to those who continued to wear their Fitbit. In addition, the predominantly well-educated, White non-Hispanic sample may limit external generalizability. The sample also had a majority of early stage breast cancer survivors and thus may not generalize to women with more advanced breast cancer. Although the initial trial was randomized, there is the potential for selection bias among those participants who decided to continue in the long-term study. Finally, knowledge of participation in the study may have conferred some effect of external accountability among participants that would not be present outside of the research setting.

Conclusions

This study examined patterns of wearable technology use and activity levels among breast cancer survivors during and after completion of a physical activity intervention. We found higher activity levels among participants receiving an intervention with greater engagement and accountability, but that activity levels reduced in follow-up after completion of the intervention. These results provide important insights regarding behavior during and after a physical activity intervention, and may help inform the design of future interventions to more effectively promote, both short- and long-term, sustained increases in physical activity.

Acknowledgments

Research support was provided by the National Cancer Institute (NCI) of the National Institutes of Health (NIH/NCI K07CA181323). HKN was supported by a Padres Pedal the Cause/Rady grant (PTC2020) and University of California San Diego Moore Cancer Center, and by a Specialized Cancer Control Support grant (NIH/NCI P30CA023100). The content in this manuscript 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.

References

1. American Cancer Society. American Cancer Society. Atlanta, GA: American Cancer Society; 2019. URL: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/cancer-treatment-and-survivorship-facts-and-figures/cancer-treatment-and-survivorship-facts-and-figures-2019-2021.pdf [accessed 2022-06-06]
2. Lahart IM, Metsios GS, Nevill AM, Carmichael AR. Physical activity, risk of death and recurrence in breast cancer survivors: A systematic review and meta-analysis of epidemiological studies. Acta Oncol 2015 May;54(5):635-654. [doi: 10.3109/0284186X.2014.998275] [Medline: 25752971]
3. Spei M, Samoli E, Bravi F, La Vecchia C, Bamia C, Benetou V. Physical activity in breast cancer survivors: A systematic review and meta-analysis on overall and breast cancer survival. The Breast 2019 Apr;44:144-152. [doi: 10.1016/j.breast.2019.02.001]

4. Schmitz KH, Campbell AM, Stuiver MM, Pinto BM, Schwartz AL, Morris GS, et al. Exercise is medicine in oncology: Engaging clinicians to help patients move through cancer. CA Cancer J Clin 2019 Nov 16;69(6):468-484 [FREE Full text] [doi: 10.3322/caac.21579] [Medline: 31617590]
5. Patel A, Friedenreich C, Moore S, Hayes S, Silver J, Campbell K, et al. American College of Sports Medicine Roundtable Report on Physical Activity, Sedentary Behavior, and Cancer Prevention and Control. Med Sci Sports Exerc 2019 Nov;51(11):2391-2402 [FREE Full text] [doi: 10.1249/MSS.0000000000002117] [Medline: 31626056]
6. Campbell KL, Winters-Stone KM, Wiskemann J, May AM, Schwartz AL, Courneya KS, et al. Exercise Guidelines for Cancer Survivors: Consensus Statement from International Multidisciplinary Roundtable. Med Sci Sports Exerc 2019 Nov;51(11):2375-2390 [FREE Full text] [doi: 10.1249/MSS.0000000000002116] [Medline: 31626055]
7. Cancer Survivors and Physical Activity. National Cancer Institute. URL: https://progressreport.cancer.gov/after/physical_activity [accessed 2022-06-06]
8. Lucas AR, Levine BJ, Avis NE. Posttreatment trajectories of physical activity in breast cancer survivors. Cancer 2017 Jul 15;123(14):2773-2780 [FREE Full text] [doi: 10.1002/cncr.30641] [Medline: 28272835]
9. Grimmett C, Corbett T, Brunet J, Shepherd J, Pinto BM, May CR, et al. Systematic review and meta-analysis of maintenance of physical activity behaviour change in cancer survivors. Int J Behav Nutr Phys Act 2019 Apr 27;16(1):37 [FREE Full text] [doi: 10.1186/s12966-019-0787-4] [Medline: 31029140]
10. Haas BK, Kimmel G, Hermanns M, Deal B. Community-Based FitSTEPS for Life Exercise Program for Persons With Cancer: 5-Year Evaluation. JOP 2012 Nov;8(6):320-324. [doi: 10.1200/jop.2012.000555]
11. Cheifetz O, Dorsay JP, MacDermid JC. Exercise facilitators and barriers following participation in a community-based exercise and education program for cancer survivors. J Exerc Rehabil 2015 Feb 28;11(1):20-29 [FREE Full text] [doi: 10.12965/jer.150183] [Medline: 25830140]
12. Courneya KS, Vardy JL, O'Callaghan CJ, Friedenreich CM, Campbell KL, Prapavessis H, et al. Effects of a Structured Exercise Program on Physical Activity and Fitness in Colon Cancer Survivors: One Year Feasibility Results from the CHALLENGE Trial. Cancer Epidemiol Biomarkers Prev 2016 Jun;25(6):969-977. [doi: 10.1158/1055-9965.EPI-15-1267] [Medline: 27197271]
13. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
14. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
15. Carver CS, Scheier MF. Control theory: a useful conceptual framework for personality-social, clinical, and health psychology. Psychol Bull 1982 Jul;92(1):111-135. [Medline: 7134324]
16. Sullivan AN, Lachman ME. Behavior Change with Fitness Technology in Sedentary Adults: A Review of the Evidence for Increasing Physical Activity. Front Public Health 2016;4:289 [FREE Full text] [doi: 10.3389/fpubh.2016.00289] [Medline: 28123997]
17. Gell NM, Grover KW, Humble M, Sexton M, Dittus K. Efficacy, feasibility, and acceptability of a novel technology-based intervention to support physical activity in cancer survivors. Support Care Cancer 2017 Apr;25(4):1291-1300. [doi: 10.1007/s00520-016-3523-5] [Medline: 27957621]
18. Lewis ZH, Lyons EJ, Jarvis JM, Baillargeon J. Using an electronic activity monitor system as an intervention modality: A systematic review. BMC Public Health 2015 Jun 24;15:585 [FREE Full text] [doi: 10.1186/s12889-015-1947-3] [Medline: 26104189]
19. Coughlin SS, Stewart J. Use of Consumer Wearable Devices to Promote Physical Activity: A Review of Health Intervention Studies. J Environ Health Sci 2016 Nov;2(6):10.15436/2378-6841.16.1123 [FREE Full text] [doi: 10.15436/2378-6841.16.1123] [Medline: 28428979]
20. Hardcastle SJ, Jiménez-Castuera R, Maxwell-Smith C, Bulsara MK, Hince D. Fitbit wear-time and patterns of activity in cancer survivors throughout a physical activity intervention and follow-up: Exploratory analysis from a randomised controlled trial. PLoS One 2020 Oct 19;15(10):e0240967 [FREE Full text] [doi: 10.1371/journal.pone.0240967] [Medline: 33075100]
21. 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]
22. Gell NM, Grover KW, Savard L, Dittus K. Outcomes of a text message, Fitbit, and coaching intervention on physical activity maintenance among cancer survivors: a randomized control pilot trial. J Cancer Surviv 2020 Feb;14(1):80-88. [doi: 10.1007/s11764-019-00831-4] [Medline: 31776849]
23. Howlett N, Trivedi D, Troop NA, Chater AM. Are physical activity interventions for healthy inactive adults effective in promoting behavior change and maintenance, and which behavior change techniques are effective? A systematic review and meta-analysis. Transl Behav Med 2019 Jan 01;9(1):147-157 [FREE Full text] [doi: 10.1093/tbm/iby010] [Medline: 29506209]

24. Hartman SJ, Nelson SH, Weiner LS. Patterns of Fitbit Use and Activity Levels Throughout a Physical Activity Intervention: Exploratory Analysis from a Randomized Controlled Trial. JMIR Mhealth Uhealth 2018 Feb 05;6(2):e29 [FREE Full text] [doi: 10.2196/mhealth.8503] [Medline: 29402761]
25. Van Blarigan EL, Chan H, Van Loon K, Kenfield SA, Chan JM, Mitchell E, et al. Self-monitoring and reminder text messages to increase physical activity in colorectal cancer survivors (Smart Pace): a pilot randomized controlled trial. BMC Cancer 2019 Mar 11;19(1):218 [FREE Full text] [doi: 10.1186/s12885-019-5427-5] [Medline: 30866859]
26. Farnell G, Barkley J. The effect of a wearable physical activity monitor (Fitbit One) on physical activity behaviour in women: A pilot study. jhse 2017;12(4):1230-1237. [doi: 10.14198/jhse.2017.124.09]
27. DiFrancisco-Donoghue J, Jung M, Stangle A, Werner WG, Zwibel H, Happel P, et al. Utilizing wearable technology to increase physical activity in future physicians: A randomized trial. Prev Med Rep 2018 Dec;12:122-127 [FREE Full text] [doi: 10.1016/j.pmedr.2018.09.004] [Medline: 30234000]
28. Shin G, Jarrahi MH, Fei Y, Karami A, Gafinowitz N, Byun A, et al. Wearable activity trackers, accuracy, adoption, acceptance and health impact: A systematic literature review. J Biomed Inform 2019 May;93:103153 [FREE Full text] [doi: 10.1016/j.jbi.2019.103153] [Medline: 30910623]
29. Shin G, Feng Y, Jarrahi MH, Gafinowitz N. Beyond novelty effect: a mixed-methods exploration into the motivation for long-term activity tracker use. JAMIA Open 2019 Apr;2(1):62-72 [FREE Full text] [doi: 10.1093/jamiaopen/ooy048] [Medline: 31984346]
30. Lindgren T, Hooper J, Fukuoka Y. Perceptions and Experiences of Women Participating in a Digital Technology-Based Physical Activity Intervention (the mPED Trial): Qualitative Study. JMIR Public Health Surveill 2019 Dec 20;5(4):e13570 [FREE Full text] [doi: 10.2196/13570] [Medline: 31859677]
31. Wu HS, Gal R, van Sleeuwen NC, Brombacher AC, IJsselsteijn WA, May AM, et al. Breast Cancer Survivors' Experiences With an Activity Tracker Integrated Into a Supervised Exercise Program: Qualitative Study. JMIR Mhealth Uhealth 2019 Feb 21;7(2):e10820 [FREE Full text] [doi: 10.2196/10820] [Medline: 30789349]
32. Thomas S, Reading J, Shephard RJ. Revision of the Physical Activity Readiness Questionnaire (PAR-Q). Can J Sport Sci 1992 Dec;17(4):338-345. [Medline: 1330274]
33. Hartman SJ, Natarajan L, Palmer BW, Parker B, Patterson RE, Sears DD. Contemp Clin Trials 2015 Nov;45(Pt B):371-376 [FREE Full text] [doi: 10.1016/j.cct.2015.09.021] [Medline: 26427563]
34. Van Domelen DR. accelerometry: Functions for Processing Minute-to-Minute Accelerometer Data. rdrr.io. URL: tinyurl.com/2p89w3a2 [accessed 2022-06-06]
35. Choi L, Liu Z, Matthews CE, Buchowski MS. Validation of accelerometer wear and nonwear time classification algorithm. Med Sci Sports Exerc 2011 Feb;43(2):357-364 [FREE Full text] [doi: 10.1249/MSS.0b013e3181ed61a3] [Medline: 20581716]
36. Kelly LA, McMillan DG, Anderson A, Fippinger M, Fillerup G, Rider J. Validity of actigraphs uniaxial and triaxial accelerometers for assessment of physical activity in adults in laboratory conditions. BMC Med Phys 2013 Nov 26;13(1):5 [FREE Full text] [doi: 10.1186/1756-6649-13-5] [Medline: 24279826]
37. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. Med Sci Sports Exerc 1998 May;30(5):777-781. [doi: 10.1097/00005768-199805000-00021] [Medline: 9588623]

Abbreviations

GAMM: generalized additive mixed effects model
MET: metabolic equivalents
MVPA: moderate-to-vigorous physical activity

Edited by L Buis; submitted 17.02.22; peer-reviewed by S De La Torre, C Memering; comments to author 06.05.22; revised version received 20.05.22; accepted 23.05.22; published 30.06.22.

Please cite as:
Hartman SJ, Chen R, Tam RM, Narayan HK, Natarajan L, Liu L
Fitbit Use and Activity Levels From Intervention to 2 Years After: Secondary Analysis of a Randomized Controlled Trial
JMIR Mhealth Uhealth 2022;10(6):e37086
URL: https://mhealth.jmir.org/2022/6/e37086
doi:10.2196/37086
PMID:35771607

Corrigenda and Addenda

Correction: Mobile Health–Based Thermometer for Monitoring Wound Healing After Endovascular Therapy in Patients With Chronic Foot Ulcer: Prospective Cohort Study

Donna Shu-Han Lin[1,2], MD; Jen-Kuang Lee[1,2,3,4,5], MD, PhD

[1]Division of Cardiology, Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan

[2]Cardiovascular Center, National Taiwan University Hospital, Taipei, Taiwan

[3]Department of Laboratory Medicine, National Taiwan University College of Medicine, Taipei, Taiwan

[4]Telehealth Center, National Taiwan University Hospital, Taipei, Taiwan

[5]Department of Internal Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan

Corresponding Author:
Jen-Kuang Lee, MD, PhD
Division of Cardiology
Department of Internal Medicine
National Taiwan University Hospital
No 7 Rd. South Chung Shan
Taipei, 100
Taiwan
Phone: 886 223123456 ext 65412
Email: b85401104@gmail.com

Related Article:

Correction of: https://mhealth.jmir.org/2021/5/e26468

Abstract

In the “Patient Selection” section in Methods, “Patients with chronic foot ulcers were consecutively enrolled from the outpatient clinic or during admission from June 2019 to December 2019” should instead be “Patients with chronic foot ulcers were consecutively enrolled from the outpatient clinic or during admission from July 2019 to December 2019”. Similarly, in the “Methods” section in the Abstract, “Patients who had a chronic foot ulcer (>3 months) and underwent endovascular therapy between June 2019 and December 2019 were included” should read “Patients who had a chronic foot ulcer (>3 months) and underwent endovascular therapy between July 2019 and December 2019 were included”. Lastly, in the “Patient Demographics and Clinical Features” section in Results, “Between June 2019 and December 2019” should instead be “Between July 2019 and December 2019”.

(JMIR Mhealth Uhealth 2022;10(6):e39749) doi:10.2196/39749

In “Mobile Health–Based Thermometer for Monitoring Wound Healing After Endovascular Therapy in Patients With Chronic Foot Ulcer: Prospective Cohort Study” (JMIR Mhealth Uhealth 2021;9(5):e26468) the authors noted three errors.

1. In the “Methods” section in the *Abstract*, “Patients who had a chronic foot ulcer (>3 months) and underwent endovascular therapy between June 2019 and December 2019 were included” has been corrected to “Patients who had a chronic foot ulcer (>3 months) and underwent endovascular therapy between July 2019 and December 2019 were included.”
2. In the “Patient Selection” section in *Methods*, “Patients with chronic foot ulcers were consecutively enrolled from the outpatient clinic or during admission from June 2019 to December 2019” has been corrected to “Patients with chronic foot ulcers were consecutively enrolled from the outpatient clinic or during admission from July 2019 to December 2019.”
3. In the “Patient Demographics and Clinical Features” section in *Results*, “Between June 2019 and December 2019” has been corrected to “Between July 2019 and December 2019.”

The correction will appear in the online version of the paper on the JMIR Publications website on June 1, 2022, 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 20.05.22; this is a non–peer-reviewed article;accepted 23.05.22; published 01.06.22.

Please cite as:
Lin DSH, Lee JK
Correction: Mobile Health–Based Thermometer for Monitoring Wound Healing After Endovascular Therapy in Patients With Chronic Foot Ulcer: Prospective Cohort Study
JMIR Mhealth Uhealth 2022;10(6):e39749
URL: https://mhealth.jmir.org/2022/6/e39749
doi:10.2196/39749
PMID:35649250

Corrigenda and Addenda

Correction: “Viewing Mobile Health Technology Design Through the Lens of Amplification Theory”

Beza Merid[1], PhD; Maria Cielito Robles[2], BSc; Brahmajee K Nallamothu[3], MD, MPH; Mark W Newman[4], PhD; Lesli E Skolarus[2], MD

[1]School for the Future of Innovation in Society, Arizona State University, Tempe, AZ, United States

[2]Department of Neurology, University of Michigan Medical School, Ann Arbor, MI, United States

[3]Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI, United States

[4]School of Information, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Beza Merid, PhD
School for the Future of Innovation in Society
Arizona State University
1120 South Cady Mall
Tempe, AZ, 85281
United States
Phone: 1 (480) 727 8787
Email: Beza.Merid@asu.edu

Related Article:

Correction of: https://mhealth.jmir.org/2022/6/e31069

(JMIR Mhealth Uhealth 2022;10(6):e40273) doi:10.2196/40273

In “Viewing Mobile Health Technology Design Through the Lens of Amplification Theory” (JMIR Mhealth Uhealth 2022;10(6):e31069) the authors have one correction.

In the original publication, the last name of author *'Maria Cielito Robles'* was listed as *'Cielito Robles'*. This has now been corrected to *'Robles.'*

This change is also reflected in the citation information of the paper as *'Robles MC.'*

The correction will appear in the online version of the paper on the JMIR Publications website on June 28, 2022, 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 15.06.22; this is a non–peer-reviewed article;accepted 15.06.22; published 28.06.22.

Please cite as:
Merid B, Robles MC, Nallamothu BK, Newman MW, Skolarus LE
Correction: “Viewing Mobile Health Technology Design Through the Lens of Amplification Theory”
JMIR Mhealth Uhealth 2022;10(6):e40273
URL: https://mhealth.jmir.org/2022/6/e40273
doi:10.2196/40273
PMID:35763796

Original Paper

Health Consumer Engagement, Enablement, and Empowerment in Smartphone-Enabled Home-Based Diagnostic Testing for Viral Infections: Mixed Methods Study

Cynthia LeRouge[1,2], MS, PhD; Polina Durneva[1], BA; Victoria Lyon[2,3], BA, MPH; Matthew Thompson[2], MD, PhD

[1]Department of Information Systems and Business Analytics, College of Business, Florida International University, Miami, FL, United States
[2]Primary Care Innovation Lab, Department of Family Medicine, University of Washington, Seattle, WA, United States
[3]Get-Grin Inc, Austin, TX, United States

Corresponding Author:
Cynthia LeRouge, MS, PhD
Department of Information Systems and Business Analytics
College of Business
Florida International University
11200 S.W. 8th St., RB 206B
Miami, FL, 33199
United States
Phone: 1 305 348 4709
Email: clerouge@fiu.edu

Abstract

Background: Health consumers are increasingly taking a more substantial role in decision-making and self-care regarding their health. A range of digital technologies is available for laypeople to find, share, and generate health-related information that supports their health care processes. There is also innovation and interest in home testing enabled by smartphone technology (smartphone-supported home testing [smart HT]). However, few studies have focused on the process from initial engagement to acting on the test results, which involves multiple decisions.

Objective: This study aimed to identify and model the key factors leading to health consumers' engagement and enablement associated with smart HT. We also explored multiple levels of health care choices resulting from health consumer empowerment and activation from smart HT use. Understanding the factors and choices associated with engagement, enablement, empowerment, and activation helps both research and practice to support the intended and optimal use of smart HT.

Methods: This study reports the findings from 2 phases of a more extensive pilot study of smart HT for viral infection. In these 2 phases, we used mixed methods (semistructured interviews and surveys) to shed light on the situated complexities of health consumers making autonomous decisions to engage with, perform, and act on smart HT, supporting the diagnostic aspects of their health care. Interview (n=31) and survey (n=282) participants underwent smart HT testing for influenza in earlier pilot phases. The survey also extended the viral infection context to include questions related to potential smart HT use for SARS-CoV-2 diagnosis.

Results: Our resulting model revealed the smart HT engagement and enablement factors, as well as choices resulting from empowerment and activation. The model included factors leading to engagement, specifically various intrinsic and extrinsic influences. Moreover, the model included various enablement factors, including the quality of smart HT and the personal capacity to perform smart HT. The model also explores various choices resulting from empowerment and activation from the perspectives of various stakeholders (public vs private) and concerning different levels of impact (personal vs distant).

Conclusions: The findings provide insight into the nuanced and complex ways health consumers make decisions to engage with and perform smart HT and how they may react to positive results in terms of public-private and personal-distant dimensions. Moreover, the study illuminates the role that providers and smart HT sources can play to better support digitally engaged health consumers in the smart HT decision process.

(JMIR Mhealth Uhealth 2022;10(6):e34685) doi:10.2196/34685

KEYWORDS

smart HT; mHealth; patient engagement; patient enablement; patient empowerment; diagnostic testing; viral infection; patient activation; consumer health informatics; influenza; home testing; mobile phone

Introduction

Emerging Smartphone-Enabled Home Testing

"If we can get a test that everyone wakes up and, just like they put in their contact lenses, they take a test, and if it turns positive, they stay at home...it will stop the vast majority of transmission and cause these outbreaks to disappear in a matter of weeks" [1].

Health consumers are increasingly taking a more substantial role in decision-making and self-care of their health [2,3]. This role includes using home-based diagnostic tests (also called home tests, self-tests, or home-use tests), where new technologies can expand and enhance our ability to "examine the body's inner workings and preoffer an exact explanation of the person's present medical condition" [4]. Home testing is convenient and enhances the efficiency of obtaining test results. Home tests are generally publicly available (eg, can be sold over the counter). They allow health consumers to obtain and test self-collected specimens from their location. Home test consumers can interpret test results independently without the help of trained health professionals [5]. Home tests differ from home collection kits (eg, 23AndMe), which require individuals to self-collect samples at home, mail them to a laboratory or clinic for analysis, and obtain the results via phone or a web-based portal. The more immediate results of home testing also potentially help to avoid the spread of infections [6]. Currently, numerous biotechnical institutions are targeting new frontiers in self-diagnostic innovations for viral infections that aim to be client centered, technically robust, and financially affordable [7-9].

Health information technology (HIT) is now seen as a fundamental aspect of patient care as it stimulates patient engagement and encourages personal health management [10]. Furthermore, health care providers increasingly demand patient interaction with digital health technologies to enroll in care, access personal health information, communicate with providers, and monitor health [11]. Coupling technology with testing supplies needed to obtain specimens (eg, tubes, containers and swabs) for home tests can support and reinforce the decision process and ultimate health care path resulting from diagnostic testing. Specifically, smartphone-supported home testing (smart HT) is receiving increasing interest and can give health consumers the ability to play a more active role in the testing experience [12-14].

Smart HT content and features support engaged health consumers in testing safely and independently in their homes, learning how to manage their illness based on test results, learning how to manage the spread, and sharing test results for personal or public health networks electronically [15,16]. Smart HT may be particularly promising to support personal and public health concerns (ie, contribute to public health surveillance and management) related to respiratory viruses, such as influenza and COVID-19. Furthermore, smart HT may leverage new convenient means of connecting to care options, potentially minimizing the spread of respiratory viruses. Specifically, a smart HT accommodating a telemedicine encounter allows enabled health consumers to act on results through an at-home connection with health providers, thereby expediting suitable personal care and minimizing contact with others when quarantine is appropriate.

Consumer health tools, including smart HT, must be effectively designed and used [10]. Therefore, it is increasingly important to understand consumer HIT patterns, including who uses specific technologies, how technologies are accessed, factors associated with their use, and perceived and actual benefits [10]. Regarding the practicalities of home testing success, there is an underlying assumption that the home-based tester is engaged in the testing process, enabled to perform the test, and empowered to act in ways conducive to their health (and the health of others) after receiving results. These assumptions involve multiple critical decisions that health consumers must make regarding acquiring the test, self-performing the diagnostic test, and choosing healthy choices and behaviors after testing (particularly in response to positive test results).

Smart HT Empowerment and Activation Journey

For infectious disease management, the goal of using smart HT is for individuals to receive test results and take the best course of action based on their test results for themselves and society. A holistic understanding of this journey is required for smart HT to positively affect both individual and public health. Indeed, feasibility cannot be genuinely achieved until health consumers intending to use smart HT are aware, engaged, and empowered and ultimately respond actively to the test results.

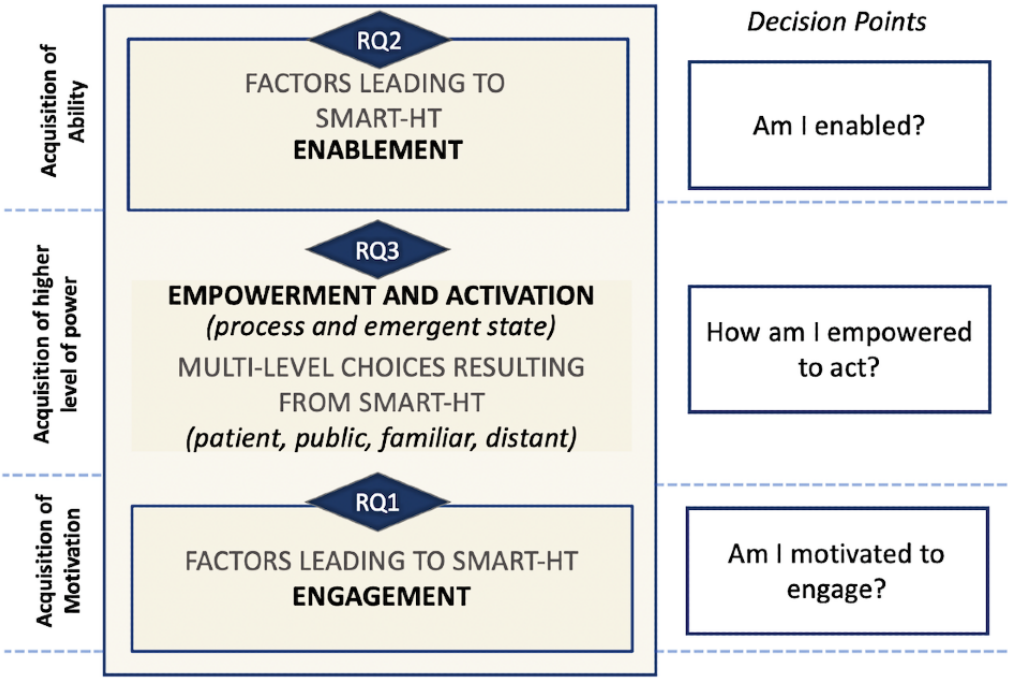
Figure 1 illustrates a patient engagement, enablement, empowerment, and activation process model (hereafter referred to as the Smart HT–Empowered Activation Model) informed by work, resulting from an extensive literature review of these states by Fumagalli et al [17]. This process model was adapted to the context of smart HT. The path to empowered activation includes healthy consumers' responses to critical personal assessments, leading to emergent states of engagement, enablement, and empowerment. Our model shows that achieving each state is ultimately based on a series of autonomous assessments in response to the following types of questions:

- Am I motivated to engage with the test? (state of engagement)
- Am I enabled to perform the test? (state of enablement)
- How am I empowered to act on the results of the test? (state of empowerment and activation)

It is important to note that this process model assumes that a health consumer is aware of and has access to smart HT. Awareness of smart HT can result from the potential user being the recipient of marketing or trial recruitment efforts (eg, through trial enrollment, marketplace, and provider) that promote the acquisition of smart HT. The factors associated

with awareness have been addressed in prior research [14,18]. Health consumers who are aware of and have access to smart HT can become engaged, enabled, empowered, and activated through smart HT use (as illustrated in Figure 1). Descriptions of these emergent states (engagement, enablement, empowerment, and activation) are presented in Textbox 1.

Figure 1. Smartphone-supported home testing (Smart HT)–Empowered Activation Model (research questions, RQ's, highlighted).



Textbox 1. Descriptions of engagement, enablement, empowerment, and activation.

Engagement

- Engagement refers to individual motivation to participate in self-management behaviors.

Enablement

- Enablement comprises 2 components:
  - having appropriate knowledge, skills, and abilities to understand one's health condition and make decisions.
  - having appropriate contexts to learn such knowledge, skills, and abilities

Empowerment and activation

- Empowerment is a consequence of enablement and engagement and takes a form of an emergent state and process:
  - As an emergent state, empowerment allows individuals to have an active role in their own care.
  - As a process, empowerment is a process of "activating" individuals, indicating that someone gains knowledge of how to manage their health condition and access appropriate health care.

Note: Descriptions derived from a literature review performed in Fumagalli et al [17].

According to Fumagalli et al [17], engagement and enablement are critical for achieving empowerment and activation. In the context of smart HT, engaged health consumers are those who develop the motivation to engage with smart HT, specifically to test for their health condition by using smart HT. However, engagement alone is not sufficient to achieve empowerment and activation, as it is also essential for health consumers to become enabled. Enabled health consumers have the appropriate knowledge of how to complete smart HT and the capacity to perform the test. In the context of smart HT, the technological aspect of the test is an important component supporting health consumers' efforts to complete the test successfully. Therefore, various characteristics of technology need to be considered when exploring enablement.

As health consumers acquire engagement and enablement, they achieve an emergent state of empowerment and activation. When viewed as an emergent state, empowered health consumers possess a higher level of power and appreciation for their role in the health care process [17]. An activated patient is "someone who has...the skills and behavioral repertoire to manage their condition, collaborate with their health providers, maintain their health functioning, and access appropriate and

high-quality care" [17]. Empowerment and activation can be a recursive process for smart HT as initial empowerment may be fueled by individual steps completed successfully as smart HT testing is enacted, which further fuels empowerment for downstream steps.

When empowerment is coupled with activation, possession of knowledge, skills, attitudes, and self-awareness can improve individuals' life situations. In the context of smart HT, health consumers become empowered to enact behaviors that could affect them personally (eg, self-care) and affect the public (eg, self-isolation to prevent spread) upon receiving positive test results.

The attainment of enablement and engagement is affected by multiple factors. The extensive literature that informed the model in Figure 1 provides some insight into the basic concepts and definitions of engagement, enablement, empowerment, and activation (Textbox 1) [17]. However, we still have a limited understanding of the factors affecting health consumers' path toward engagement and enablement and health consumer choices resulting from empowerment and activation.

HIT studies that address the antecedents of consumer health technology use [19-21] do not generally distinguish the factors related to moving toward states of engagement and enablement. Instead, these HIT studies tend to focus on demographic factors (eg, race, sex, and socioeconomic status), health conditions (eg, overweight or obese), or adherence to healthy behaviors (eg, eating or physical activity patterns) holistically affecting adoption without recognizing the emergent states in the process leading to use or acceptance [21-25]. Furthermore, the literature on HIT adoption does not explore the various pathways for smart HT–empowered action. Therefore, in the case of smart HT, we know little about consumers' choice of options and intentions once enabled by the test results. Thus, to fully uncover and understand consumer patterns, it is imperative to understand the factors influencing the enablement and engagement states and the movement to the empowerment and activation process and emergent states.

This study aimed to gain a deeper understanding of the process related to achieving patient activation for smart HT by understanding the factors that affect decisions to move along the empowered activation process. We addressed this exploration in the context of respiratory viral infection (RVI), which is a serious public health threat [26], meriting smart HT exploration and consideration. We specifically targeted the 3 decision points, particularly interactions with smart HT, by addressing the following research questions (RQs):

- RQ1: For a health consumer aware of smart HT for RVIs, what factors lead to the emergent state of engagement with smart HT?
- RQ2: For a health consumer aware of smart HT for RVIs, what factors lead to the emergent state of enablement to perform smart HT?
- RQ3: For a health consumer who is enabled and engaged with smart HT for RVIs, what choices result from the emergent state of empowerment to act upon the results (particularly positive test results) obtained through smart HT?

Methods

Study Overview

The focus of this project is a pilot study of an innovation called flu@home, a smart HT for influenza. This flu@home pilot is part of a more extensive research study called the Seattle Flu Study, which explored the feasibility of using home-based testing for the surveillance and public health management of viral outbreaks [27,28]. The flu@home smart HT contains 2 major components: a mobile app and an influenza test kit. The mobile app was designed to screen participants experiencing influenza-like illness (ILI) symptoms and facilitate testing of participants. Once screened, the participants used the app to consent to the research protocol and order their influenza test kit. The influenza test kit included materials adapted from an existing point-of-care lateral flow test called the QuickVue Influenza A+B test (Quidel Corporation). Once participants received the influenza test kit, the mobile app gave them instructions to complete the self-test.

The flu@home pilot comprised four phases: (1) flu@home smart HT usability study, (2) trial of flu@home, (3) semistructured interviews regarding the experience of using flu@home, and (4) a survey of those who used flu@home. Figure 2 summarizes the 4 phases of the study and describes the objectives of each phase.

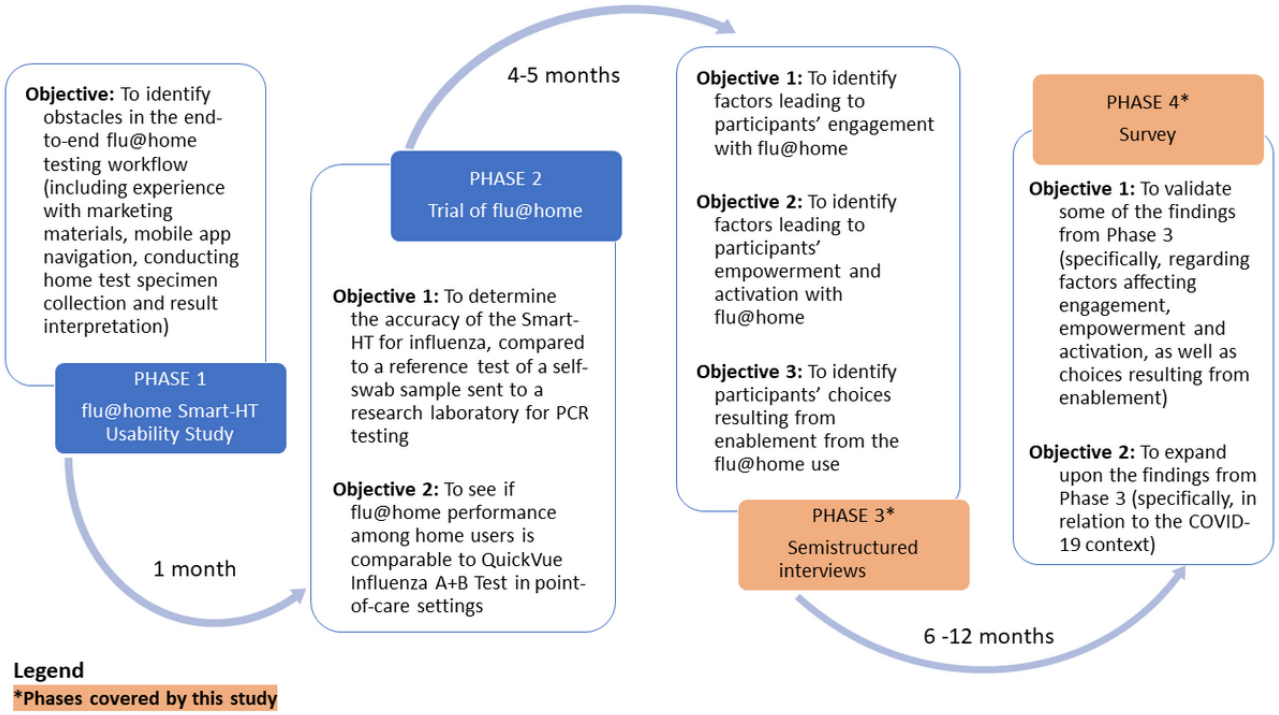
Phase 1 (the flu@home smart HT usability study) focused on the development of flu@home to meet usability standards. Participants from phase 1 usability assessments used to inform the software development were not recruited for the subsequent phases.

After the development of flu@home, phase 2 (trial of flu@home) was conducted to determine its accuracy. During phase 2, the participants had a chance to experience the actual flu@home test. Phase 2 participants were also recruited for phases 3 and 4, which explored participants' experiences with the flu@home test, various factors affecting engagement and enablement, and choices resulting from empowerment and activation.

These multiple phases of the flu@home pilot leveraged mixed methods (both qualitative and quantitative). Mixed methods can be valuable for developing and evaluating complex interventions such as smart HT [29,30]. Studies have recognized that mixed methods add value by identifying the mechanisms of complex problems, increasing the validity of the findings, and providing a deeper understanding of the phenomenon of interest [31,32]. In this study, we used an exploratory sequential design described by Creswell and Clark [33], which was used first to explore a phenomenon of interest (through qualitative methods) and then clarify the findings by leveraging quantitative methods. In line with this approach, we collected qualitative data (phase 3) to explore the decision points and factors associated with engagement, enablement, and empowerment. We then conducted a quantitative phase (phase 4) to validate and further explore various factors affecting decision points associated with engagement and enablement and choices resulting from empowerment and activation.

The inclusion criteria for phase 2, which also applied to phases 3 and 4, involved eligible participants who were aged ≥18 years, spoke English, had an iPhone or iPad, and had an ILI (defined as the presence of a cough and at least one or more of the following symptoms: fever, chills or sweats, muscle, body aches, or feeling tired or more tired than usual). Recruitment was limited to individuals in the lower 48 states of the United States to ensure that they received their flu@home test kit within 2 days of enrolling in the study. Overall, 97.9% (724/739) of participants who completed phase 2 consented to be contacted for future, related research and were eligible to participate in phases 3 and 4.

Figure 2. Study phases (phases 3 and 4 [shown in orange] are the objectives that relate to the scope of this paper). PCR: Polymerase chain reaction.



Semistructured Interviews

Participants from phase 2 (trial of flu@home) were invited to participate in in-depth, semistructured interviews to share their experiences with the flu@home smart HT and their beliefs and attitudes toward using smart HT (for influenza) in the future. Semistructured interviews involved a series of predetermined, open-ended questions with probes and prompts to elicit further information about the phenomenon of interest [34]. We used a phenomenological approach to conduct semistructured interviews. Phenomenology allows researchers to explore human experiences to elicit meanings for individuals by analyzing their perceptions of the phenomenon of interest [35]. In particular, we leveraged hermeneutic phenomenology [36]. In hermeneutic phenomenology (as distinguished from transcendental phenomenology), pre-existing knowledge and researchers' understanding of concepts related to the phenomenon of interest cannot be fully bracketed in interpreting participants' descriptions of the phenomenon [37]. Phenomenological hermeneutic semistructured interviews were leveraged to gain insights into factors influencing decision points associated with engagement, empowerment, and activation, as well as choices resulting from enablement [38-41].

Our interview guide (Multimedia Appendix 1 [18,42-46]) aligns with the phenomenological interview method described by Bevan [47]. The interview guide included a series of broad and open-ended questions that allowed participants to express their opinions extensively and freely. We recognized three aspects in developing a phenomenological interview guide: (1) contextualization (understanding of participants' context in which the experience of the phenomenon of interest is situated), (2) apprehending the phenomenon (questions related to the specific phenomenon of interest), and (3) clarifying the phenomenon (eg, imaginative variation) [47]. In alignment with the principle of contextualization, our interviews started with general questions related to participants' general attitudes and behaviors related to health. After discussing participants' general attitudes and behaviors related to health, our interview questions transitioned to exploring participants' experiences with flu@home and the values and gains associated with using flu@home (ie, apprehending the phenomenon). Although not specifically referencing our high-level constructs of interest in the discussion, these questions aimed to explore engagement, enablement, and activation in more detail. Finally, we clarified the smart HT phenomenon using imaginative variation [47]. Imaginative variation is leveraged when the researcher understands a specific element of a participant's experience, which is then applied to varying its structural components to uncover the invariant parts. We used imaginative variation to explore hypothetical situations, such as using smart HT in the future for influenza and other medical issues.

Prior literature was reviewed to inform the semistructured interview questions and a priori coding schema for data analysis. We looked to the literature that would provide more insight and detail into elements of the conceptual framework used in this

study, precisely, factors affecting engagement and enablement and choices resulting from empowerment and activation (Figure 1). Two studies that included extensive literature review especially informed our interview content and a priori coding schema: the Fumagalli Concept Map of Engagement and Neighboring Concepts [17] and the Digital Health Engagement Model (DIEGO) [18].

Fumagalli et al [17] indicated that engagement manifests in patients' behaviors to improve their role in health care. Patient motivation to engage in such behaviors can be determined by intrinsic influences (eg, proactive role in health care resulting in the patient making appointments, staying informed about treatment options, and others) and extrinsic influences (eg, specific characteristics of health intervention) [17,48]. In alignment with this view of engagement, we focused on 2 categories of factors affecting decision points associated with engagement: intrinsic and extrinsic influences.

To inform categories of factors associated with enablement, we referred to applicable high-level concepts in the DIEGO [18] The DIEGO model contains multiple categories of factors associated with an individual's enrollment in and engagement with digital health interventions. Some categories of factors in the DIEGO model point to specific aspects of health consumers' interactions with digital health, which can enable health consumers to complete a digital health intervention. We leveraged factors particularly pertinent to health consumers' ability to complete the intervention: considering the quality (of the HIT) and assessing personal capacity (for using the HIT).

Finally, to inform the interview questions and high-level codes for categories of choices resulting from health consumers' empowerment and activation, we considered the different levels of impact (*patient* and *public*). These different levels of impact were partially informed by the DIEGO, which examined individual-level and public-level engagement with digital health interventions [18]. We subdivided these categories to consider proximal associations (familiar and distant). Overall, 4 multilevel choices emerged: *patient-familiar*, *patient-distant*, *public-familiar*, and *public-distant*. *Patient-familiar* actions are defined as actions that individuals take to care for their illness in a familiar setting (eg, visiting primary care providers or self-managing the illness). *Patient-distant* actions are ways in which individuals can seek care in a more distant manner (eg, visiting urgent care or seeking telemedicine consultations). *Public-familiar* actions are actions that individuals take to prevent the spread of their illness to family, friends, coworkers, and people they interact with frequently. Finally, *public-distant* actions are those that individuals take to prevent the spread of their illness in their community at large, such as sharing their test results to contribute to the awareness of the illness in their community.

The interview sample size was guided by data saturation, which is the point at which additional data collection no longer generates any new insights [49]. Prior studies with similar study designs indicate that data saturation can generally be achieved in data samples ranging from 10 to 40 individuals [50-53]. Therefore, in alignment with prior studies and general recommendations for sample sizes, we determined a minimum of 20 interviews to be an appropriate target number.

We recruited participants in 3 waves to include a diverse representation of geographic locations and ages (to accurately reflect the targeted user population). The first wave of the selection process comprised sorting participants into age groups (18-24, 25-34, 35-44, 45-64, and ≥65 years) and randomly inviting them to interviews, selecting participants from each group. During this initial recruitment wave, we sent 60 participants study invitations assuming that 50% of participants would sign up for an interview based on completion rates of home collection studies for other health conditions [54-56]. In the 2 subsequent waves of recruitment, the proportion of participants recruited from each age group was adjusted to ensure sample representation across all age groups. Recruitment continued until at least 3 participants from each age group were interviewed in each stratum. A total of 115 participants were invited, and 31 (26.9%) completed the interviews. Table 1 summarizes the participants' demographics in phase 2 (trial of flu@home) and phase 3 (specifically, participants who were invited to participate in the semistructured interviews and participants who completed the semistructured interviews).

Confidential 40- to 60-minute semistructured interviews were conducted using Zoom videoconferencing [57]. A total of 3 research team members with backgrounds in HIT, consumer technologies, and public health conducted the interviews. In cases where 2 research team members were present, 1 team member served as the lead interviewer. The other team members served the role of scribe and active listeners. The 2 team members conducting the interview held a debriefing session after each interview to discuss key points to consider for coding purposes and discuss the interview protocol flow. Deidentified interview transcripts were uploaded to Dedoose, a software for qualitative data analysis.

Thematic analysis was conducted to code the deidentified interview transcripts. Thematic analysis allows the identification, analysis, description, and reporting of themes found in qualitative data [58]. We established the validity and reliability of the thematic analysis results by following the Lincoln and Guba [52] criteria for conducting qualitative research. (researcher triangulation, code reviews, expert feedback, and resolution meetings [52,59]).

First, we inductively coded (ie, created low-level codes) our interview transcripts without referring to our conceptual model (Figure 1). Inductive coding allowed us to capture phenomenological user experiences with flu@home. Second, after inductive coding, we referred to an a priori high-level coding schema (Textbox 2) reflective of our conceptual model (Figure 1). In particular, we reviewed our low-level codes to determine potential connections with high-level concepts (ie, engagement, enablement, empowerment, and activation). During this step, we found conceptual associations between low-level codes and high-level concepts in the model.

Table 1. Demographic data of the sample frame (phase 2 participants) used for semistructured interview (phase 3).

Phase (sample size)	Phase 2: trial of flu@home (n=724), n (%)	Phase 3: invited to participate in semistructured interviews (n=115), n (%)	Phase 3: completed semistructured interviews (n=31), n (%)
Age (years)			
18-24	86 (11.9)	12 (10.4)	3 (9.7)
25-34	204 (28.2)	34 (29.6)	6 (19.4)
35-44	199 (27.5)	38 (33)	11 (35.5)
45-64	188 (25.9)	21 (18.3)	8 (25.8)
≥65	47 (6.5)	10 (8.7)	3 (9.7)
Ethnicity			
White	510 (70.4)	78 (67.8)	21 (67.7)
Black or African American	63 (8.7)	10 (8.7)	6 (19.4)
Asian	60 (8.3)	8 (6.9)	0 (0)
Native Hawaiian or other Pacific Islander	4 (0.6)	1 (0.9)	1 (3.2)
American Indian or Alaska Native	17 (2.4)	18 (15.7)	1 (3.2)
N/A[a], other, or prefer not to say	70 (9.7)	2 (1.7)	2 (6.5)
Geographic representation			
West	214 (29.6)	43 (37.4)	14 (45.2)
Midwest	139 (19.2)	21 (18.3)	5 (16.1)
Southwest	12 (1.7)	2 (1.7)	1 (3.2)
Northeast	197 (27.2)	32 (27.8)	4 (12.9)
Southeast	162 (22.4)	17 (14.8)	7 (22.6)

[a]N/A: not applicable.

Textbox 2. Categories of factors (engagement; enablement) and choices (empowerment and activation).

Engagement

- Intrinsic influences
- Extrinsic influences

Enablement

- Considering the quality
- Assessing personal capacity

Empowerment and activation

- Patient-familiar
- Patient-distant
- Public-familiar
- Public-distant

To enhance research reliability and validity, the research team used a constant comparison method to refine coding [49,60]. This procedure involved 2 coders (unfamiliar with the conceptual model) and an internal auditor (a research team member with extensive qualitative research methods expertise who was familiar with the conceptual model). The internal auditor also reviewed the structure, syntax, and labeling of the final coding schema and performed a code review of 100% of the coded quotes to ensure alignment with the final coding structure. The coding team met regularly to iteratively discuss and reconcile initial inductive coding, which included ensuring that codes were supported by linked quotes, refining coding categories, and reviewing emerging themes. Once a detailed inductive coding scheme was in place, the coding team independently and collectively identified and reconciled the conceptual associations between the low-level codes (created

because of inductive coding) and high-level concepts (Textbox 2). The team members traced the codes forward from code to model and backward from model to detailed codes and their underlying quotes from the transcripts. Throughout both inductive and deductive coding, the coding team collectively discussed and resolved any identified issues with codes associated with supporting quotes, as well as the structure, syntax, and labeling of the final comprehensive model. Reconciling points mainly focused on combining various subcodes and updating the labels and definitions of individual codes.

Surveys

The survey contained 3 sections. The first section contained questions about the participants' prior engagement with the smart HT for influenza. The second section contained questions about the impact of the COVID-19 pandemic on participants' future engagement and enablement decisions associated with smart HT for influenza. Finally, the third section contained questions about participants' potential engagement, enablement, empowerment, and activation decisions associated with smart HT for COVID-19. The survey did not ask participants to provide demographic information, given institutional review board cautions in asking demographics to preserve the anonymous nature of the survey (thus, we were unable to perform an analysis of demographic and categorical data).

Insights from semistructured interviews informed the survey questions, which were developed to validate and further explore factors affecting decision points associated with engagement and enablement, as well as choices resulting from empowerment and activation. Further exploration of factors was conducted because of the emergence of the COVID-19 pandemic. The research team included additional questions related to engagement, empowerment, and enablement associated with smart HT for COVID-19. The survey questions used ordinal and categorical response options. For ordinal questions, the research team used a 5-item Likert scale for responses ranging from strongly agree to strongly disagree. The Likert scale is an efficient and reliable technique for examining individual attitudes and perceptions [61,62]. Compared with 7- or 10-point scales, 5-point Likert scales have been shown to reduce survey fatigue and increase response rates [63,64]. In addition to Likert-type questions, the survey included a few categorical responses that aligned with the nature and purpose of the questions and were not well-suited to a Likert scale. The research team reviewed all the survey questions to ensure clarity. The results are shared in Multimedia Appendix 2, and the *Results* section of this paper showcases the survey questions relevant to this study.

Participants from phase 2 (trial of flu@home) were recruited to complete the survey (Table 1 shows participants' demographics from phase 2). The participants received an initial recruitment email and a follow-up email a week later. They did not receive compensation for completing the survey, and no demographic information was collected.

The anonymous survey was administered using Qualtrics Survey Platform software [65] in June 2020. In total, 38.2% (282/739) of eligible individuals from phase 2 completed the survey.

Survey data were analyzed using descriptive statistics (mode, as well as response distribution counts and percentages), as appropriate for Likert scales [66].

Ethics Approval

The study design was approved by the University of Washington Institutional Review Board (STUDY00007627).

Results

Smart HT Engagement Overview

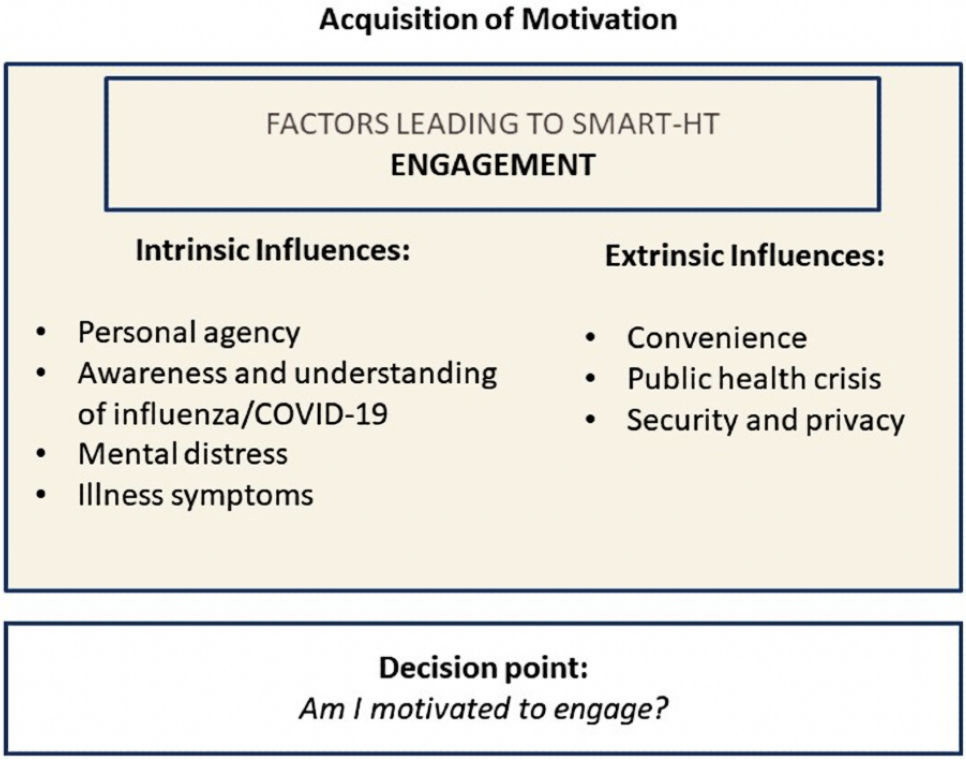
We aligned the general structure of our results with the Smart HT–Empowered Activation Model (Figure 1), which showcases the emergent states of engagement, enablement, empowerment, and activation covered by this study. We provide associated key interviews and survey highlights under the associated Smart HT–Empowered Activation Model sections. We provide further details of our findings in Multimedia Appendix 3 (interview evidence trace table) and Multimedia Appendix 2 (details of ordinal survey response questions).

Acquisition of Motivation: Smart HT Engagement

Overview

The acquisition of motivation involved both intrinsic and extrinsic influences. We identified intrinsic influences covering specific states (eg, mental distress) and traits (eg, personal agency) of the users. In addition, the identified extrinsic influences covered specific characteristics of smart HT (eg, convenience) and environmental conditions (eg, public health crisis). Figure 3 summarizes these findings.

Figure 3. Factors affecting engagement. Smart HT: smartphone-supported home testing.



Smart HT Engagement: Intrinsic Influence Factors

Intrinsic influences include personal agency, awareness and understanding of viral infection (influenza; COVID-19), mental distress, and illness symptoms.

Personal agency was stated as a contributing factor to engaging with smart HT. The participants generally expressed high personal agency in managing their health. Most interview participants articulated that they believed they were primarily responsible for managing their health or acted as equal partners in their care with their providers. Interviewees perceived the flu@home smart HT as providing them with choice and control over when and where to conduct diagnostic testing.

This technology appealed to interview participants who proactively managed their health, as well as individuals who self-identified as having poor health behaviors. Therefore, although we recognize that this may play a role in the decision process for some, there was no general consensus that *health behaviors and attitudes* were influential factors in participants' considerations of engaging with smart HT.

Awareness and understanding of the illness (in this study, influenza and COVID-19) was a factor in the interview participants' consideration of engaging with smart HT. Interviewees' beliefs regarding the severity of seasonal influenza varied greatly. Those who believed influenza was a minor illness (frequently referencing a *cold*) were less motivated to engage with a self-test than those who perceived influenza as a serious health concern.

Regarding when health consumers might be motivated to perform a smart HT, survey responses indicated (Multimedia Appendix 2) that 96.5% (272/282) of participants strongly or somewhat agreed that they would use the flu@home test if they experienced ILI *symptoms*. Only 32.6% (92/282) of participants strongly agreed that they would use flu@home testing when asymptomatic. As for COVID-19 testing, 95.8% (249/260) of the participants strongly agreed or somewhat agreed that, if available, they would use a COVID-19 home test when experiencing COVID-19–related symptoms. Most survey respondents (196/260, 75.4% strongly agreed or somewhat agreed) stated that they would use a COVID-19 home test, even if they did not experience symptoms common to COVID-19. To further understand the relationship between the acquisition of information and motivation stages, it is notable that symptom onset did not necessarily correlate with the preferred timing of test acquisition. Specifically, some interview participants indicated that they might opt to proactively purchase smart HT to keep at home for convenient access when symptoms (and the need for testing) arise. One of the participants explained the following:

> *I should keep some of the kits at home on an ongoing basis so that anytime I feel I have this fever, sneezing, runny nose and all those symptoms.*

Moreover, interview participants also indicated that mitigating *mental distress* at a time when they were also feeling physically unwell was a motivating factor for smart HT. When asked what would make them inclined to use a smart HT instead of going to a physician, one participant said the following:

> *And I think that it would be more convenient because sometimes you just don't feel well and feel like leaving the house. It'd be nice because I feel like you might be able to find out results sooner than if you wait to go to the doctor...I feel like there's less pressure when you're at home, and you're more relaxed...I shouldn't say pressure, but less stress. There's always some, at*

> *least for me, levels of extra stress going to the doctor just in general. Getting out of the house and sitting in the waiting room, and being back there, and just kind of like nervous and stuff, waiting to see what the doctor's going say.*

We further delved into this issue in the survey. Although testing at home when feeling ill may mitigate the stress of traveling and waiting to see a provider to perform a diagnostic test, a concern about the future of home testing is that test results indicating a serious health condition (such as COVID-19) delivered without provider support could cause mental distress. According to our survey results, 74.1% (192/259) of the survey respondents indicated that they would find testing positive (ie, learning that they contracted the COVID-19 illness) in the at-home context (smart HT) no more distressing than learning of their diagnosis in other health care settings (Multimedia Appendix 4).

Smart HT Engagement: Extrinsic Influence Factors

Extrinsic influences include convenience, public health crises, and security and privacy. Participants overwhelmingly shared that *convenience* was a primary motivating factor in considering smart HT use. For example, one of the participants stated the following:

> *Just the idea of being able to do home-based checking interests me. It sounds like it has promise to me. And I think that a lot of people might use something like that rather than going through the grief of trying to get a doctor's appointment, which is hard to do here.*

Convenience of engaging with smart HT manifested in avoiding the burden of visiting a provider in person. Participants mentioned some of the burdens of visiting a provider, including difficulty in scheduling appointments, finding appointments that would not require taking time off from work, and difficulty meeting a provider in person while caring for young children. Some interview participants indicated that they were particularly motivated to use smart HT to diagnose their children.

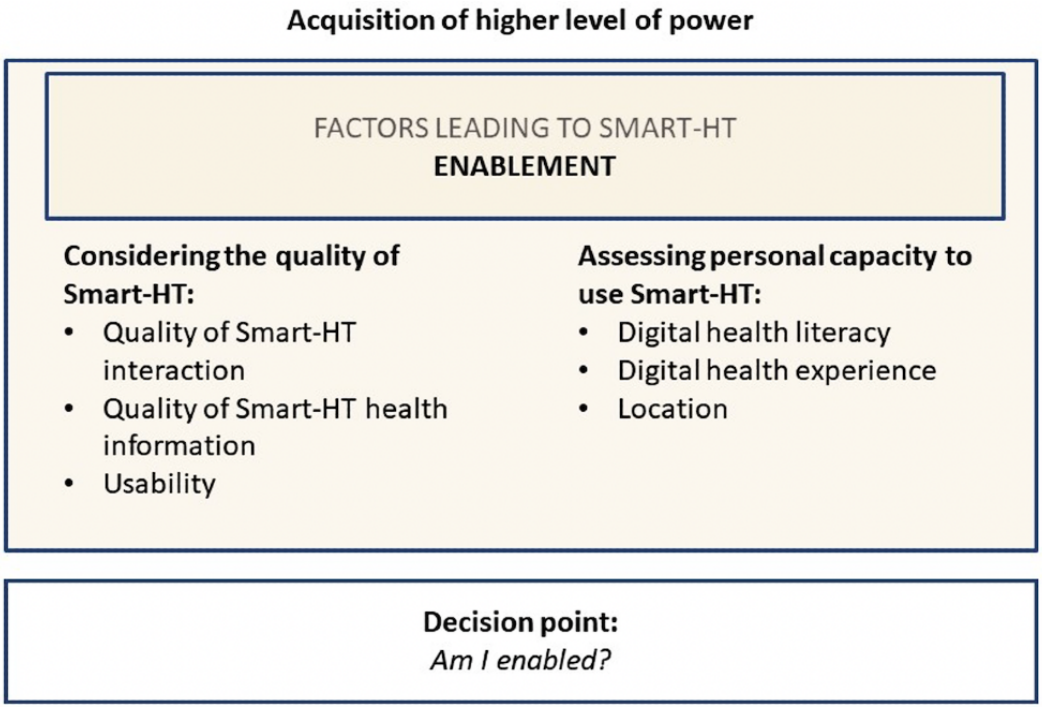
Furthermore, survey respondents indicated that a *public health crisis* (ie, the COVID-19 pandemic) affected their decision to engage with smart HT for viral infections. Participants' attitudes toward smart HT were affected by the COVID-19 pandemic (Multimedia Appendix 2). For example, when asked, "Which of the following best describes how COVID-19 influences your thoughts about using flu@home to test for common, seasonal flu if you have symptoms?" the most common response (132/263, 50.2%) was, "I am much more likely to use flu@home for testing of common, seasonal flu."

Interview participants also identified the *security and privacy* of the data generated from smart HT as a factor of engagement. Participants were most concerned that their health data would be sold or shared without their consent if tests were provided via web-based sources.

Acquisition of Ability: Enablement Through Smart HT

The interview data revealed various factors leading to participants' enablement facilitated by smart HT. We categorized these factors into two groups: (1) considering the quality of smart HT and (2) assessing personality capacity to use smart HT (Figure 4).

Figure 4. Factors affecting enablement. Smart HT: smartphone-supported home testing.



Enablement: Considering the Quality of Smart HT

Regarding the *quality of smart HT interactions*, interview participants shared that they valued a digital experience that segmented the testing process into small, digestible, step-by-step instructions with illustrations and videos in the app. The participants also appreciated the built-in timers to reduce the likelihood of errors.

Interviewees also considered the *quality of smart HT health information.* Participants indicated that they understood and retained content that differentiated common cold and influenza symptoms, as well as information provided when it was appropriate to consult a provider. Participants were also receptive to the flu@home app, including general facts about influenza, such as how many people are affected each year, and other information to help remind consumers about the importance of preventive measures (such as vaccination).

Participants also shared that they would like future iterations of flu@home to include more health information, such as explaining "how contagious the flu is," and recommendations for managing the illness. For example, one of the participants described the following:

> *I think one of the biggest things we do that we shouldn't do in this day and age is just trying to take something [medication] to suppress the symptoms and then head right back to work or other things. I think having some statistics or data about the dangers of taking the flu out of the house...would be really helpful.*

In evaluating the *usability* of the testing processes, interview participants described the flu@home system as easy to use, attributable to clearly labeled kit contents and simple nasal swabbing procedures with mild or no discomfort. These features seemed to influence their decision to perform the test. Incorporating a smartphone app into home testing broadens the potential features that test developers can incorporate into smart HT. The study team presented many potential features in interviews with participants to consider enhanced value. Added features that appealed to the participants included the ability to share their home test results with their providers. Many participants were interested in smart HT that incorporated data collected from a wearable device. In addition, participants were interested in receiving alerts if an RVI outbreak occurred in or near their community. Nearly all participants indicated that they would be willing to share deidentified data to contribute to their community's public health management of influenza. However, participants' responses varied greatly regarding whether they would value gamification features in smart HT.

Enablement: Assessing Personal Capacity to Use Smart HT

The interview participants generally indicated that they felt capable of completing the smart HT test. Specifically, participants indicated that they felt enabled to complete the test with the digital guidance provided in the flu@home app, thus, informing their belief that they completed the swabbing procedures as instructed in the app and that they felt capable of completing the smart HT again in the future.

Digital health literacy was a critical factor in the assessment of personal capacity. All participants indicated that they could download the app, order the kit, and complete the testing without clinician supervision. They demonstrated that they had the skills to complete these actions because they completed the pilot study in phase 1. However, frequent *digital health experience* (eg, using mobile apps and wearables) was not a universal factor in participants' assessment of their capacity to use smart HT. Although some participants described their lifestyle as including the use of a wearable device or health app, others said they had not found such products to be valuable and did not use digital health resources unless necessary for clinical care.

Moreover, the *location* of the testing was another critical factor in assessing the personal capacity to perform the test. Participants generally indicated that finding a specific place in their home was essential for performing the test. Participants mentioned various locations where they could perform the test (eg, bathroom, bedroom, and kitchen) and the specific characteristics of such locations. For instance, some participants indicated that it was critical for the location where they performed the test to be clean. In addition, one participant alluded to privacy as an essential aspect of choosing a location for the test. This participant said the following:

> *I feel like I would do it at home because there's no other people around. They wouldn't just see me stick something, the little test tube up my nose, or whatever. Can't even think of what it's called.*

Acquisition of Higher Level of Power: Empowerment and Activation

Empowerment and Activation Overview

In the case of smart HT, empowerment and activation involve 2 sequential points. The first factor was the intention to perform the test. The second was the intention to act on confirmation of influenza results from completing the smart HT test.

Empowerment and Activation: Intention to Test Using Smart HT

We found evidence that the study population was empowered to use smart HT to test for viruses. Approximately 81.5% (207/254) of the participants who completed the survey indicated that they would prefer to test for viruses at home rather than a test conducted by a health care provider (Multimedia Appendix 4). Moreover, survey results indicated that 94.6% (265/280) of participants somewhat to strongly agreed that they would use the smart HT test kit for influenza in the future, regardless of pandemic conditions (see Multimedia Appendix 2 for details). Analogously, interview participants shared that they intended to acquire and use smart HT in the future, once commercially available.

Survey responses showed that many people were willing to test for COVID-19 regularly, every 14 days (97/260, 37.3%), or monthly (93/260, 35.8%) to ensure that they were healthy and could interact with others (Multimedia Appendix 4).

Moreover, there were indications that an empowerment and activation process could have spillover effects on other possibilities. During the interviews, participants also mentioned

various other (not necessarily viral) health conditions that they would be interested in using smart HT in the future: common cold, dementia, bronchitis, cancer, diabetes, pneumonia, sinus infections, hepatitis, and many others.

Empowerment and Activation: Intention to Act on Confirming Influenza Results

The findings indicate that individuals consider actions related to all four themes regarding acting on positive smart HT results:(1) patient-familiar, (2) patient-distant, (3) public-familiar, and (4) public-distant. Figure 5 summarizes these findings.

Textbox 3 summarizes the qualitative results that support multilevel choices resulting from empowerment and activation.

In looking holistically at multilevel choices resulting from empowerment and activation, some interview participants gravitated toward *patient-familiar* means of self-care (eg, self-management or primary care provider appointment). However, other interview participants were open to less familiar forms of care, such as urgent care or emergency room visits or seeking a telemedicine (virtual) consultation *(patient-distant).* Limited access to care for reasons such as rural living status and insurance coverage were mentioned in the interviews as deciding factors for self-management of illness or high motivation for a virtual consultation. In addition, the rationale shared for virtual consultation included convenience and treatment expedience (eg, antiviral prescription), potentially minimizing the chance of spreading the illness and acquiring a new illness during a provider visit.

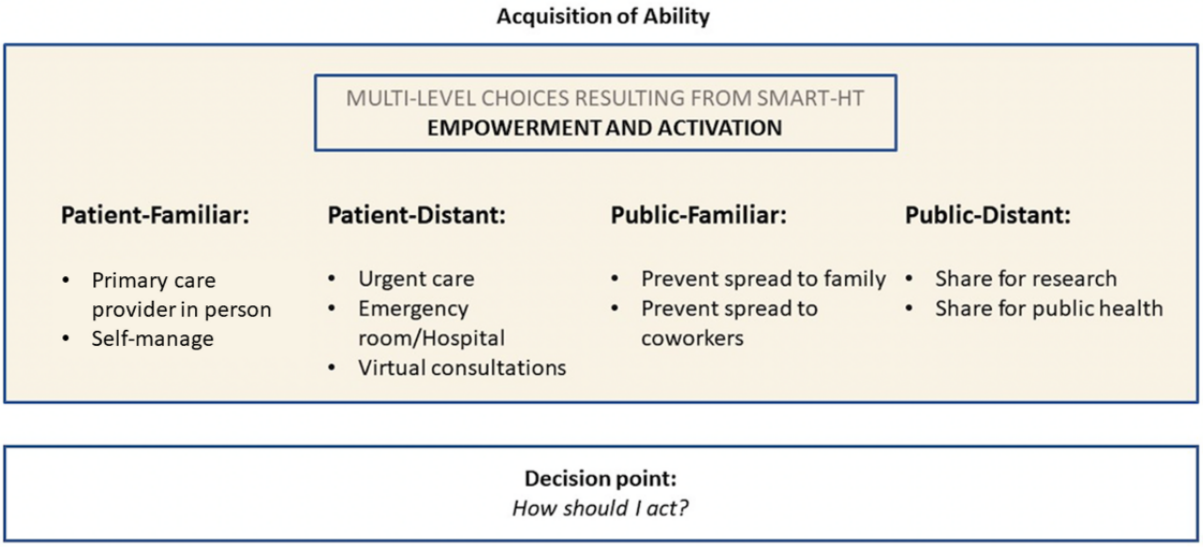
To assess whether participants' attitudes toward telemedicine (virtual care) changed because of COVID-19, we asked them to reflect on their initial willingness to seek virtual care. Survey respondents indicated that they were equally willing to have a virtual care appointment (telemedicine) after testing positive for influenza and COVID-19. Approximately 93.9% (265/282) of the participants strongly or somewhat agreed that they would have been willing to have a virtual appointment if the flu@home results had returned positive. Similarly, 93.1% (242/260) of the participants somewhat or strongly agreed that they would have been willing to have a virtual appointment if their COVID-19 test results returned positive.

It is also noteworthy that interview participants reported that the responsibility of caring for young children influenced their test result response choices, with parents of young children sometimes opting for distance care for themselves but preferring in-person care for their children.

Regarding public considerations, interview participants indicated that they were receptive to contributing to the public health management of a viral outbreak (*public-familiar)*. In addition to the *public-familiar* means of managing the spread provided in Multimedia Appendix 4 for influenza, most survey respondents indicated that they were taking some of the recommended actions to prevent the spread of COVID-19.

Regarding the *public-distant* choice, participants indicated that they were willing to share data for research purposes. Although the contribution to research generally denotes a distal relationship, participants indicated that they were more likely to participate in research studies if they were familiar with the research organization (trusting the entity to secure their data and maintain confidentiality). Moreover, most participants indicated that the COVID-19 pandemic influenced their motivation to share their smart HT results anonymously for public surveillance or research purposes. The interview data also seemed to indicate an escalated motivation for parents to contribute to the community or public health management of influenza.

Figure 5. Multilevel choices resulting from empowerment and activation. Smart HT: smartphone-supported home testing.



Textbox 3. Key findings from interviews related to multilevel choices resulting from empowerment and activation.

Patient-familiar

- Self-manage: Reasons for self-management included assumptions that a provider would tell them to rest at home and take over-the-counter medication to manage their illness anyway and the cost of care (particularly for the uninsured). Motivations for moving from self-care included the perceived need for prescription medication.
- Primary care provider in person: Participants indicating that they would seek an appointment with their provider frequently referenced an established, trusting relationship with their primary care provider: some referenced pre-existing conditions that could create health care complexities with influenza.

Patient-distant

- Urgent care: Rationale for urgent care as a form of provider engagement included reference to accessibility, namely, urgent care clinic during weekend or evening hours.
- Hospital emergency room: Some participants indicated that they would seek care from the emergency room as their default option when unsure how to manage a health issue.
- Virtual consultation: Virtual consultation was referenced as a means of convenient verification of diagnosis and a quick means to obtain treatment (ie, prescriptions).

Public-familiar

- Prevent spread to family: Although participants referenced both quarantining in and sanitizing their homes to prevent spread to family, they also shared practical challenges, particularly with quarantine.
- Prevent spread to coworkers: Participants mentioned the preference to stay home when sick with influenza symptoms to prevent spread to coworkers and more distant relationships (eg, public transit commuters). There was also mention of practical challenges because of some work arrangements.

Public—distant

- Share for research: Some participants felt that anonymously sharing self-test results could contribute to improved influenza vaccine development.
- Share for public health: Participants were generally willing to share self-test results for surveillance if done anonymously. They also indicated that they would personally reference a local or neighborhood-level influenza map in making prevention choices.

Discussion

Principal Findings

Smart HT aspires to facilitate the success and impact of home-based diagnostic testing by coupling the diagnostic procedure with technological supports [12-14]. Our findings implicitly signal promising aspects of coupling home-based medical procedures with digital support. However, for smart HT to achieve its intended use advantages, health consumers must have increasing levels of awareness, motivation, and ability to perform home-based diagnostic tests and act on results appropriately. Our mixed methods study results provide insight into the nuanced and myriad factors that affect engagement and enablement with smart HT for viral infections, as well as how empowered users intend to respond to smart HT results. Overall, this study extends a stream of past work by exploring each of these concepts (patient engagement, empowerment, activation, and enablement) in the smart HT context [17]. Our final Smart HT–Empowered Activation Model may have implications for an increased understanding of engagement, enablement, and empowerment in other HIT contexts.

Essentially, we contribute to the existing knowledge by *opening the black box* of engagement, enablement, and empowerment by contextualizing these constructs in the context of smart HT for viral infections. We were guided by the Fumagalli Concept Map of Engagement and Neighboring Concepts [17] and the DIEGO [18]. These models are based on an extensive literature review that depicts a health consumer's progression toward the acquisition of a higher level of power (g, appreciating one's role in health care and managing one's health) that allows them to directly participate in the care process. Our results identify factors (from the perspective of the health consumer) that come into play for the engagement, enablement, and empowerment emergent states and frame our findings in the Smart HT–Empowered Activation Model.

Our findings highlight the complexity of digital health engagement. One of the most apparent elements of complexity is the number of factors that come into play during the empowered activation journey. For smart HT information technology developers, this indicates the importance of having a strategy to consider, leverage, and support various factors that lead to successfully performing the test and acting on test results, essentially, a journey map (as noted in design thinking [67]) that showcases the potential of the technology to support the test process.

Upon further reviewing our model through the lens of complexity, it is noteworthy that various factors involving the acquisition of motivation (engagement) can change over time. For example, regarding intrinsic factors, changing illness symptoms and awareness and understanding of RVIs can affect motivation to engage with smart HT. This timing element is something to consider, particularly in the role of technology in message engagement considerations. Public health crises, an extrinsic example, can influence an individual's motivation to

use smart HT. Smart HT technology features can help health consumers evaluate the safety and feasibility of testing at home versus seeking testing by other means during public health crises. This motivating factor begets developers and perhaps policy makers to promote the use of smart HT with demonstrated efficacy during a public health crisis.

Regarding intrinsic motivation, smart HT technology can provide information to enhance a user's awareness and understanding of a specific respiratory illness. Providing this educational information also introduces some assessment regarding how much information is provided, when to provide the information, and whether information needs to be tailored to a particular user's capabilities, base knowledge, or interest in the information. Additionally, it is important to note that some factors of engagement may be more important than others. For example, there were mixed findings regarding the importance of participants' general health behaviors and attitudes. Some participants indicated that they generally practiced healthy behaviors, whereas others indicated the opposite. This factor may be subordinate to other factors (eg, personal agency and illness symptoms). The relative importance of these factors should be explored in future research.

Furthermore, regarding the acquisition of motivation (engagement), we found evidence that individuals did not anticipate feeling more distress when learning that they have an RVI at home using smart HT than when learning about their health status in a clinical setting. However, research indicates that diagnosing different health conditions such as cancer, dementia, and COVID-19 can evoke emotional distress [68-70]. Limited research assessing the mental distress of home testing exists. As more home-based diagnostic tools are developed and available for consumer use for various health conditions, future research is needed to understand the mental distress of receiving different types of diagnoses through a home-based test compared with clinical settings. Anticipating and mitigating mental anguish because of a positive test result may be worth considering in the design and use of smart HT. Pretest and posttest counseling have been suggested in some forms of home testing [71]. When relevant, smart HT technology features and functions may either provide functions to mitigate distress directly or refer the user to resources for assistance in managing distress.

The study shows that in the case of contagious diseases, multiple level factors need to be considered to have a robust smart HT. The acquisition of a higher level of power (empowerment and activation) involves decisions at multiple levels, which can have both personal and far-reaching impacts. For example, informed individuals may vary in their patient, familiar, distant, and public actions when testing positive for a viral infection. Ideally, this choice variance is because of an informed decision process and not because of missing information or misinformation. Therefore, a key role of smart HT during the acquisition of ability (enablement) stage is to prepare the individual performing the test for the multiple downstream choices resulting from enablement. In response, developers may want to embed quality information and various paths of action into the design and functions of smart HT to support an informed empowerment and activation decision-making process. The multiple choices presented to a patient upon receipt of their positive test results should be carefully considered when developers design smart HTs to reduce choice complexity. It is particularly important to ensure that patients are not overwhelmed with too many choices, as too many options can impair an individual's subsequent self-control (and, therefore, personal agency) [72]. Furthermore, choices should be limited to *good choices*, indicating choices that align with the overall purpose of smart HT. For instance, if a patient tests positive for influenza, smart HT can provide them with a set of suggestions on managing their conditions and preventing their spread to other people. In addition, there can be an option for digitally sharing the results for research or public health purposes with smart HT. Moreover, smart HT might provide easy access to a digital provider (telemedicine) for treatment. To further minimize complexity and guarantee choices that keep the smart HT user heading in the direction in which they want to go, developers should try to limit recommendations or choices to those tailored to individual smart HT users. In addition, to facilitate empowerment regarding tailored choices, smart HT should provide clear guidance regarding the next steps required and the use of information for each possible option.

In addition to providing a better understanding of the affirmative path of the Smart HT–Empowered Activation Model, our work provides a foundation for future work to explore other paths through the model. For example, future empirical research could explore the relative importance of the identified factors or whether the factors would hold in the context of other forms of smart HT. Regarding the latter, participants did mention potential interest in various forms of smart HT for both acute and chronic conditions. Future research could explore and validate the factors of engagement, empowerment, and enablement (derived in this study) in other contexts mentioned by the participants.

Limitations

As with most studies that include qualitative methods, the generalization of the results must be approached with some caution. Although the study included a diverse set of participants in terms of age and geography, it had some limitations. Most notably, our study population included only participants in the flu@home study. Attitudes toward home testing may differ among individuals who did not experience this specific smart HT or type of respiratory testing. In addition, this study was conducted early during the COVID-19 pandemic. COVID-19 home testing and self-swabbing availability and experiences during the pandemic might have implications for smart HT for influenza and other home-based diagnostic testing. Overall, we strongly encourage future research to consider our findings in other smart HT and HIT contexts.

Conclusions

Through our findings, we proposed and informed a Smart HT–Empowered Activation Model depicting an engagement, enablement, empowerment, and activation process for smart HT use. The resulting model underscores the need to understand and address the path to health consumer empowerment and activation for smart HT use, resulting in actions that provide maximum health benefits to individuals and society. Overall,

this study provides a foundation for researchers and developers to explore and create successful engagement strategies to align with consumer digital health opportunities to promote prevention, self-care, and spread control of infectious viruses such as influenza.

Acknowledgments

The Seattle Flu Study was funded by Gates Ventures. The funder was not involved in the design of the study and does not have any ownership over the management and conduct of the study, data, or rights to publish. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Seattle Flu Study Investigators.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview protocol.
[DOCX File , 128 KB - mhealth_v10i6e34685_app1.docx]

Multimedia Appendix 2
Likert scale survey questions.
[DOCX File , 20 KB - mhealth_v10i6e34685_app2.docx]

Multimedia Appendix 3
Evidence trace tables for analysis of interview data.
[DOCX File , 20 KB - mhealth_v10i6e34685_app3.docx]

Multimedia Appendix 4
Survey results.
[DOCX File , 34 KB - mhealth_v10i6e34685_app4.docx]

References

1. Nania R. Simple over-the-counter home tests needed to slow COVID-19 spread. American Association of Retired Persons. URL: https://www.aarp.org/health/conditions-treatments/info-2020/home-testing-coronavirus.html [accessed 2021-04-29]
2. Petrakaki D, Hilberg E, Waring J. Between empowerment and self-discipline: governing patients' conduct through technological self-care. Soc Sci Med 2018 Sep;213:146-153 [FREE Full text] [doi: 10.1016/j.socscimed.2018.07.043] [Medline: 30081356]
3. Sharma AE, Rivadeneira NA, Barr-Walker J, Stern RJ, Johnson AK, Sarkar U. Patient engagement in health care safety: an overview of mixed-quality evidence. Health Aff (Millwood) 2018 Nov;37(11):1813-1820 [FREE Full text] [doi: 10.1377/hlthaff.2018.0716] [Medline: 30395509]
4. Kearns AJ, O'Mathúna DP, Scott PA. Diagnostic self-testing: autonomous choices and relational responsibilities. Bioethics 2010 May;24(4):199-207. [doi: 10.1111/j.1467-8519.2008.00711.x] [Medline: 19222443]
5. Ibitoye M, Frasca T, Giguere R, Carballo-Diéguez A. Home testing past, present and future: lessons learned and implications for HIV home tests. AIDS Behav 2014 May 27;18(5):933-949 [FREE Full text] [doi: 10.1007/s10461-013-0668-9] [Medline: 24281697]
6. Procop GW, Kadkhoda K, Rhoads DD, Gordon SG, Reddy AJ. Home testing for COVID-19: benefits and limitations. Cleve Clin J Med 2021 Mar 01 [FREE Full text] [doi: 10.3949/ccjm.88a.ccc071] [Medline: 33579779]
7. Abbott's Fast, $5, 15-minute, easy-to-use COVID-19 Antigen Test receives FDA emergency use authorization; mobile app displays test results to help our return to daily life; ramping production to 50 million tests a month. Abbott Park. URL: https://tinyurl.com/52ahxa8f [accessed 2020-05-29]
8. Berger K. COVID-19 at-home test kits: what we know. Single Care. URL: https://www.singlecare.com/blog/news/at-home-coronavirus-tests/ [accessed 2020-07-25]
9. Coronavirus (COVID-19) Update: FDA authorizes first COVID-19 test for self-testing at home. U.S. Food & Drug Administration. URL: https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-authorizes-first-covid-19-test-self-testing-home [accessed 2022-06-09]
10. Hung M, Conrad J, Hon SD, Cheng C, Franklin JD, Tang P. Uncovering patterns of technology use in consumer health informatics. Wiley Interdiscip Rev Comput Stat 2013 Nov 31;5(6):432-447 [FREE Full text] [doi: 10.1002/wics.1276] [Medline: 24904713]

11. Guendelman S, Broderick A, Mlo H, Gemmill A, Lindeman D. Listening to communities: mixed-method study of the engagement of disadvantaged mothers and pregnant women with digital health technologies. J Med Internet Res 2017 Jul 05;19(7):e240 [FREE Full text] [doi: 10.2196/jmir.7736] [Medline: 28679489]
12. Abbott's New NAVICA App: what you need to know. Abbott. URL: https://www.abbott.com/corpnewsroom/diagnostics-testing/abbotts-new-NAVICA-app-what-you-need-to-know.html [accessed 2021-04-11]
13. Acculi Labs develops mobile app for detection of COVID 19 infected individuals. PharmaBiz. URL: http://pharmabiz.com/NewsDetails.aspx?aid=129938&sid=2 [accessed 2021-04-10]
14. Dharanikota S, LeRouge CM, Lyon V, Durneva P, Thompson M. Identifying enablers of participant engagement in clinical trials of consumer health technologies: qualitative study of influenza home testing. J Med Internet Res 2021 Sep 14;23(9):e26869 [FREE Full text] [doi: 10.2196/26869] [Medline: 34519664]
15. Xiao M, Tian F, Liu X, Zhou Q, Pan J, Luo Z, et al. Virus detection: from state-of-the-art laboratories to smartphone-based point-of-care testing. Adv Sci (Weinh) 2022 Apr 07:e2105904 [FREE Full text] [doi: 10.1002/advs.202105904] [Medline: 35393791]
16. Wood CS, Thomas MR, Budd J, Mashamba-Thompson TP, Herbst K, Pillay D, et al. Taking connected mobile-health diagnostics of infectious diseases to the field. Nature 2019 Feb 27;566(7745):467-474 [FREE Full text] [doi: 10.1038/s41586-019-0956-2] [Medline: 30814711]
17. Fumagalli LP, Radaelli G, Lettieri E, Bertele' P, Masella C. Patient Empowerment and its neighbours: clarifying the boundaries and their mutual relationships. Health Policy 2015 Mar;119(3):384-394. [doi: 10.1016/j.healthpol.2014.10.017] [Medline: 25467286]
18. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
19. Chau K, Lam M, Cheung M, Tso E, Flint S, Broom D, et al. Smart technology for healthcare: exploring the antecedents of adoption intention of healthcare wearable technology. Health Psychol Res 2019 Mar 11;7(1):8099-8009 [FREE Full text] [doi: 10.4081/hpr.2019.8099] [Medline: 31583292]
20. Laxman K, Krishnan S, Dhillon J. Barriers to adoption of consumer health informatics applications for health self management. Health Sci J 2015;9(5):1-7 [FREE Full text] [doi: 10.1201/B13617-39]
21. Deng Z, Hong Z, Ren C, Zhang W, Xiang F. What predicts patients' adoption intention toward mHealth services in China: empirical study. JMIR Mhealth Uhealth 2018 Aug 29;6(8):e172 [FREE Full text] [doi: 10.2196/mhealth.9316] [Medline: 30158101]
22. Drake C, Cannady M, Howley K, Shea C, Snyderman R. An evaluation of mHealth adoption and health self-management in emerging adulthood. AMIA Annu Symp Proc 2019;2019:1021-1030 [FREE Full text] [Medline: 32308899]
23. Hung M, Jen W. The adoption of mobile health management services: an empirical study. J Med Syst 2012 Jun 29;36(3):1381-1388. [doi: 10.1007/s10916-010-9600-2] [Medline: 20878452]
24. Ndayizigamiye P, Kante M, Shingwenyana S. An adoption model of mHealth applications that promote physical activity. Cogent Psychol 2020 May 12;7(1):1764703 [FREE Full text] [doi: 10.1080/23311908.2020.1764703]
25. Zhang X, Guo X, Lai K, Guo F, Li C. Understanding gender differences in m-health adoption: a modified theory of reasoned action model. Telemed J E Health 2014 Jan;20(1):39-46. [doi: 10.1089/tmj.2013.0092] [Medline: 24161004]
26. Carvajal L, Perret PC. Epidemiology of respiratory infections. In: Pediatric Respiratory Diseases. Cham: Springer; 2020.
27. Chu H, Boeckh M, Englund J, Famulare M, Lutz B, Nickerson D, et al. The Seattle Flu Study: a multiarm community-based prospective study protocol for assessing influenza prevalence, transmission and genomic epidemiology. BMJ Open 2020 Oct 07;10(10):e037295 [FREE Full text] [doi: 10.1136/bmjopen-2020-037295] [Medline: 33033018]
28. Lyon V, Zigman Suchsland M, Chilver M, Stocks N, Lutz B, Su P, et al. Diagnostic accuracy of an app-guided, self-administered test for influenza among individuals presenting to general practice with influenza-like illness: study protocol. BMJ Open 2020 Nov 19;10(11):e036298 [FREE Full text] [doi: 10.1136/bmjopen-2019-036298] [Medline: 33444172]
29. Homer C, Klatka K, Romm D, Kuhlthau K, Bloom S, Newacheck P, et al. A review of the evidence for the medical home for children with special health care needs. Pediatrics 2008 Oct;122(4):e922-e937. [doi: 10.1542/peds.2007-3762] [Medline: 18829788]
30. Nutting PA, Miller WL, Crabtree BF, Jaen CR, Stewart EE, Stange KC. Initial lessons from the first national demonstration project on practice transformation to a patient-centered medical home. Ann Fam Med 2009;7(3):254-260 [FREE Full text] [doi: 10.1370/afm.1002] [Medline: 19433844]
31. McKim CA. The value of mixed methods research. J Mixed Methods Res 2016 Jul 08;11(2):202-222. [doi: 10.1177/1558689815607096]
32. Hansen M, O'Brien K, Meckler G, Chang AM, Guise J. Understanding the value of mixed methods research: the Children's Safety Initiative-Emergency Medical Services. Emerg Med J 2016 Jul 23;33(7):489-494 [FREE Full text] [doi: 10.1136/emermed-2015-205277] [Medline: 26949970]
33. Creswell J, Plano Clark V. Designing and Conducting Mixed Methods Research. Los Angeles: SAGE; 2011.

34. Ayres L. Semi-structured interview. In: The SAGE Encyclopedia of Qualitative Research Methods. Thousand Oaks, California: SAGE Publications; 2008.
35. Creswell J, Poth C. Qualitative Inquiry and Research Design Choosing Among Five Approaches. Thousand Oaks, California: SAGE Publications; 2016.
36. Alase A. The interpretative phenomenological analysis (IPA): a guide to a good qualitative research approach. IJELS 2017 Apr 30;5(2):9. [doi: 10.7575/aiac.ijels.v.5n.2p.9]
37. Patton M. Qualitative Research and Evaluation Methods. Thousand Oaks, California: SAGE Publications; 1980.
38. Lewin S, Glenton C, Oxman AD. Use of qualitative methods alongside randomised controlled trials of complex healthcare interventions: methodological study. BMJ 2009 Sep 10;339(sep10 1):b3496 [FREE Full text] [doi: 10.1136/bmj.b3496] [Medline: 19744976]
39. Bunniss S, Kelly D. Research paradigms in medical education research. Med Educ 2010 Apr;44(4):358-366. [doi: 10.1111/j.1365-2923.2009.03611.x] [Medline: 20444071]
40. Creswell J. The Political in Qualitative Methods Qualitative Inquiry and Research Design: Choosing among Five Traditions. Thousand Oaks: Thousand Oaks: Sage Publications; 1998.
41. Flowers P, Larkin M, Smith J. Interpretative Phenomenological Analysis Theory, Method and Research. Thousand Oaks, California: SAGE; 2009.
42. Bartlett G, Macgibbon B, Rubinowicz A, Nease C, Dawes M, Tamblyn R. The importance of relevance: willingness to share eHealth data for family medicine research. Front Public Health 2018;6:255 [FREE Full text] [doi: 10.3389/fpubh.2018.00255] [Medline: 30234095]
43. Ben-Zeev D, Brenner CJ, Begale M, Duffecy J, Mohr DC, Mueser KT. Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. Schizophr Bull 2014 Nov;40(6):1244-1253 [FREE Full text] [doi: 10.1093/schbul/sbu033] [Medline: 24609454]
44. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
45. Chinn D, McCarthy C. All Aspects of Health Literacy Scale (AAHLS): developing a tool to measure functional, communicative and critical health literacy in primary healthcare settings. Patient Educ Couns 2013 Feb;90(2):247-253. [doi: 10.1016/j.pec.2012.10.019] [Medline: 23206659]
46. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
47. Bevan MT. A method of phenomenological interviewing. Qual Health Res 2014 Jan 10;24(1):136-144. [doi: 10.1177/1049732313519710] [Medline: 24413767]
48. Rodriguez KM. Intrinsic and extrinsic factors affecting patient engagement in diabetes self-management: perspectives of a certified diabetes educator. Clin Ther 2013 Feb;35(2):170-178. [doi: 10.1016/j.clinthera.2013.01.002] [Medline: 23411000]
49. Glaser BG. The constant comparative method of qualitative analysis. Social Problem 1965 Apr;12(4):436-445. [doi: 10.2307/798843]
50. Sim J, Saunders B, Waterfield J, Kingstone T. Can sample size in qualitative research be determined a priori? Int J Soc Res Methodol 2018 Mar 27;21(5):619-634. [doi: 10.1080/13645579.2018.1454643]
51. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
52. Lincoln YS, Guba EG. But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. New Directions Program Eval 1986;1986(30):73-84. [doi: 10.1002/ev.1427]
53. Onwuegbuzie AJ, Leech NL. Validity and qualitative research: an oxymoron? Qual Quant 2006 May 25;41(2):233-249. [doi: 10.1007/s11135-006-9000-3]
54. Carozzi B, Barrett S. An audit of home testing kits. In: Proceedings of the Joint Annual Conference of the British Association for Sexual Health and HIV & the Society for the Study of Sexually Transmitted Diseases. 2017 Presented at: The Joint Annual Conference of the British Association for Sexual Health and HIV & the Society for the Study of Sexually Transmitted Diseases; Jun 18-20, 2017; Ireland URL: https://www.morressier.com/o/event/58ef894fd462b80290b503ea/article/58ef94ddd462b80290b50542 [doi: 10.26226/morressier.58ef94ddd462b80290b50542]
55. Manavi K, Hodson J. Observational study of factors associated with return of home sampling kits for sexually transmitted infections requested online in the UK. BMJ Open 2017 Oct 22;7(10):e017978 [FREE Full text] [doi: 10.1136/bmjopen-2017-017978] [Medline: 29061625]
56. Seguin M, Dodds C, Mugweni E, McDaid L, Flowers P, Wayal S, et al. Self-sampling kits to increase HIV testing among black Africans in the UK: the HAUS mixed-methods study. Health Technol Assess 2018 Apr;22(22):1-158 [FREE Full text] [doi: 10.3310/hta22220] [Medline: 29717978]
57. Archibald MM, Ambagtsheer RC, Casey MG, Lawless M. Using zoom videoconferencing for qualitative data collection: perceptions and experiences of researchers and participants. Int J Qual Methods 2019 Sep 11;18:160940691987459. [doi: 10.1177/1609406919874596]

58. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis. Int J Qual Methods 2017 Oct 02;16(1):160940691773384. [doi: 10.1177/1609406917733847]
59. Forero R, Nahidi S, De Costa J, Mohsin M, Fitzgerald G, Gibson N, et al. Application of four-dimension criteria to assess rigour of qualitative research in emergency medicine. BMC Health Serv Res 2018 Feb 17;18(1):120 [FREE Full text] [doi: 10.1186/s12913-018-2915-2] [Medline: 29454350]
60. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov 01;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
61. Ho GW. Examining perceptions and attitudes. West J Nurs Res 2017 May 24;39(5):674-689 [FREE Full text] [doi: 10.1177/0193945916661302] [Medline: 27456460]
62. Leung S. A comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-point Likert scales. J Soc Serv Res 2011 Jul;37(4):412-421. [doi: 10.1080/01488376.2011.580697]
63. Babakus E, Mangold W. Adapting the SERVQUAL scale to hospital services: an empirical investigation. Health Serv Res 1992 Feb;26(6):767-786 [FREE Full text] [Medline: 1737708]
64. Sachdev SB, Verma HB. Relative importance of service quality dimension: a multisectoral study. J Serv Res 2004;1:93-116 [FREE Full text]
65. Qualtrics XM. Qualtrics. 2020. URL: https://www.qualtrics.com/ [accessed 2020-02-01]
66. Robertson J. Likert-type scales, statistical methods, and effect sizes. Commun ACM 2012 May;55(5):6-7. [doi: 10.1145/2160718.2160721]
67. Dorst K. The core of 'design thinking' and its application. Design Stud 2011 Nov;32(6):521-532. [doi: 10.1016/j.destud.2011.07.006]
68. Aminzadeh F, Byszewski A, Molnar FJ, Eisner M. Emotional impact of dementia diagnosis: exploring persons with dementia and caregivers' perspectives. Aging Ment Health 2007 May;11(3):281-290. [doi: 10.1080/13607860600963695] [Medline: 17558579]
69. Cook S, Salmon P, Hayes G, Byrne A, Fisher PL. Predictors of emotional distress a year or more after diagnosis of cancer: a systematic review of the literature. Psychooncology 2018 Mar;27(3):791-801 [FREE Full text] [doi: 10.1002/pon.4601] [Medline: 29318702]
70. Sahoo S, Mehra A, Dua D, Suri V, Malhotra P, Yaddanapudi LN, et al. Psychological experience of patients admitted with SARS-CoV-2 infection. Asian J Psychiatr 2020 Dec;54:102355 [FREE Full text] [doi: 10.1016/j.ajp.2020.102355] [Medline: 33271684]
71. Ng OT, Chow AL, Lee VJ, Chen MI, Win MK, Tan HH, et al. Accuracy and user-acceptability of HIV self-testing using an oral fluid-based HIV rapid test. PLoS One 2012 Sep 17;7(9):e45168 [FREE Full text] [doi: 10.1371/journal.pone.0045168] [Medline: 23028822]
72. Vohs K, Baumeister R, Schmeichel B, Twenge J, Nelson N, Tice D. Making choices impairs subsequent self-control. In: Self-Regulation and Self-Control. Milton Park, Abingdon-on-Thames, Oxfordshire, England, UK: Routledge; 2018.

Abbreviations

DIEGO: Digital Health Engagement Model
HIT: health information technology
ILI: influenza-like illness
RQ: research question
RVI: respiratory viral infection
smart HT: smartphone-supported home testing

Edited by L Buis; submitted 07.03.22; peer-reviewed by L Marcial, T Risling; comments to author 08.04.22; revised version received 15.05.22; accepted 17.05.22; published 30.06.22.

Please cite as:
LeRouge C, Durneva P, Lyon V, Thompson M
Health Consumer Engagement, Enablement, and Empowerment in Smartphone-Enabled Home-Based Diagnostic Testing for Viral Infections: Mixed Methods Study
JMIR Mhealth Uhealth 2022;10(6):e34685
URL: https://mhealth.jmir.org/2022/6/e34685
doi:10.2196/34685
PMID:35771605

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org