JMIR mHealth and uHealth

Impact Factor (2022): 5.0
Volume 9 (2021), Issue 8 ISSN 2291-5222 Editor in Chief: Laurie Buis, PhD, MSI

Contents

Reviews

Original Papers

Viewpoints

Review

Domains and Methods Used to Assess Home Telemonitoring Scalability: Systematic Review

Salome Azevedo[1,2,3*], MSc; Teresa Cipriano Rodrigues[3*], MSc, PhD; Ana Rita Londral[1,2*], MSc, PhD

[1]Value for Health CoLAB, NOVA Medical School, Lisbon, Portugal

[2]Comprehensive Health Research Centre, NOVA Medical School, UNL, Lisbon, Portugal

[3]CEG-IST, Centre for Management Studies of Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal

[*]all authors contributed equally

Corresponding Author:
Salome Azevedo, MSc
Value for Health CoLAB
NOVA Medical School
Rua do Instituto Bacteriológico, no 5, 5-A e 5-B
Lisbon, 1099-085
Portugal
Phone: 351 218803110
Email: salome.azevedo@vohcolab.org

Abstract

Background: The COVID-19 pandemic catalyzed the adoption of home telemonitoring to cope with social distancing challenges. Recent research on home telemonitoring demonstrated benefits concerning the capacity, patient empowerment, and treatment commitment of health care systems. Moreover, for some diseases, it revealed significant improvement in clinical outcomes. Nevertheless, when policy makers and practitioners decide whether to scale-up a technology-based health intervention from a research study to mainstream care delivery, it is essential to assess other relevant domains, such as its feasibility to be expanded under real-world conditions. Therefore, scalability assessment is critical, and it encompasses multiple domains to ensure population-wide access to the benefits of the growing technological potential for home telemonitoring services in health care.

Objective: This systematic review aims to identify the domains and methods used in peer-reviewed research studies that assess the scalability of home telemonitoring–based interventions under real-world conditions.

Methods: The authors followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) guidelines and used multiple databases (PubMed, Scopus, Web of Science, and EconLit). An integrative synthesis of the eligible studies was conducted to better explore each intervention and summarize relevant information concerning the target audience, intervention duration and setting, and type of technology. Each study design was classified based on the strength of its evidence. Lastly, the authors conducted narrative and thematic analyses to identify the domains, and qualitative and quantitative methods used to support scalability assessment.

Results: This review evaluated 13 articles focusing on the potential of scaling up a home telemonitoring intervention. Most of the studies considered the following domains relevant for scalability assessment: problem (13), intervention (12), effectiveness (13), and costs and benefits (10). Although cost-effectiveness was the most common evaluation method, the authors identified seven additional cost analysis methods to evaluate the costs. Other domains were less considered, such as the sociopolitical context (2), workforce (4), and technological infrastructure (3). Researchers used different methodological approaches to assess the effectiveness, costs and benefits, fidelity, and acceptability.

Conclusions: This systematic review suggests that when assessing scalability, researchers select the domains specifically related to the intervention while ignoring others related to the contextual, technological, and environmental factors, which are also relevant. Additionally, studies report using different methods to evaluate the same domain, which makes comparison difficult. Future work should address research on the minimum required domains to assess the scalability of remote telemonitoring services and suggest methods that allow comparison among studies to provide better support to decision makers during large-scale implementation.

(JMIR Mhealth Uhealth 2021;9(8):e29381) doi:10.2196/29381

KEYWORDS

telemonitoring; scalability; home telecare; systematic review

Introduction

The Universal Health Coverage commitment aligned with the emergence of COVID-19 reinforced the value of telemedicine services and elected these services crucial to coping with the pandemic's challenges in the health care sector. Since the pandemic reached the western countries, home telemonitoring offered an alternative to control the health status of infected nonsevere patients at their homes to avoid unnecessary visits to the hospital [1].

During the early part of 2020, from a social perspective, the fast-track solution to prevent the spread of COVID-19 focused on social distancing [2]. Governments forced people to stay at home, canceled mass gatherings, imposed teleworking, and closed all educational institutions [3]. From a health care perspective, governments took some extreme measures to increase the capacity to cope with the virus, namely reduction or deferral of nonurgent care and hands-on visits, and postponement of nonurgent surgeries [4]. These measures exposed high-risk groups, such as the elderly at home, people at long-term care facilities, patients with chronic conditions, and hidden diseases [5]. Inevitably, physicians started following-up with their patients through video calls and remote monitoring to continue treatment and avoid long-term complications [6]. In parallel, health care providers launched new telehealth services to assist patients in their homes [7]. Policy makers and practitioners did not have enough information to decide which pilot intervention they should disseminate into real-world settings, considering different financial reimbursement strategies, health care system organizations, and workforce acceptance levels [8].

With technological progression and decreasing equipment costs, remote patient monitoring emerged as a telemedicine application. It comprises interactive and noninteractive technologies to support health care and monitor patients' health status in their homes [9].

Home telemonitoring is one type of remote patient monitoring, which has shown and is showing potential to improve clinical and patient-reported outcomes and ensure cost reductions for health care practices [10]. In this work, the authors consider the definition given by Paré and colleagues [11] for home telemonitoring. A service based on home telemonitoring consists of health care professionals monitoring the patient's health status at a distance. Patients or caregivers transmit their health-related data to a responsible health care professional through information and telecommunication technologies. Research on home telemonitoring showed benefits concerning health care systems' capacity constraints [12], patient empowerment, and treatment commitment [13]. It revealed significant improvement in clinical outcomes even in some diseases [11]. Despite the considerable investment in accelerating health information technology [14], there is not enough information on determining whether home telemonitoring is appropriate and feasible for implementation in a real-world context [15]. Scaling up a health intervention requires wise and efficient spending of resources [16]. Therefore, it is crucial to assess the suitability of scaling up home telemonitoring interventions with proven efficacy to provide answers to the following two questions [17]: *Does it work in practice? Is it worth it?*

To answer these questions and decide which technology-based health intervention can be scaled up for mainstream care delivery, one must assess its scalability (ie, the ability to be expanded to real-world conditions without compromising on effectiveness and access to the eligible population) [18].

Most of the studies focus only on assessing the effectiveness and costs of a health intervention. Nevertheless, these are two of many considerations to address when evaluating the potential of scaling up an intervention [19]. Other domains such as the feasibility and adaptability of the health intervention and the political or strategic contexts are rarely analyzed. As emphasized by Milat and his colleagues [15] in their recently proposed Intervention Scalability Assessment Tool (ISAT), assessing a health intervention's scalability involves considering multiple domains, such as the political and strategic contexts, workforce, and infrastructure, among others.

There is a need to conduct evidence-based studies that assess pilot interventions' potential to achieve population-wide benefits [20]. Scalability studies that also consider the intervention's suitability to the socioeconomic context in question are important to estimate the success of deploying these interventions in different contexts [15].

Owing to the lack of research on scalability analysis, in this paper, the authors present a systematic review, based on Milat and colleagues' domains [15], to identify and characterize methods used to assess the potential to scale-up home telemonitoring interventions in the context of a growing telehealth service in the industry. This study focuses on peer-reviewed studies conducted to evaluate the scalability of follow-up interventions based on home telemonitoring. The authors aim to provide a comprehensive overview of these studies concerning the domains and methods used and identify gaps for future research to address when evaluating the potential to implement or scale-up home telemonitoring interventions. As the authors are not aware of other systematic reviews focusing on this aspect, they believe that this review will enlighten researchers, practitioners, and policy makers regarding the most used strategies to assess the scalability of home telemonitoring interventions.

Methods

The search strategy followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) guidelines to conduct the review [21]. The population, intervention, comparison, outcome (PICO) framework [21] allowed the identification of key concepts such as "Home Telemonitoring," "Follow-up," "Scalability," and "Assessment" to formulate a well-focused question and facilitate the literature search. To

optimize the search through effective queries, the authors used PubMed's Medical Subject Headings (MeSH) to identify indexed terms [22]. This step was fundamental as this review emerges from the combination of research fields with different terms for the same concept. Textbox 1 presents the rationale used to build the final query used in each database.

Textbox 1. Queries used to search each database.

1. (((Telemonitoring) OR (Home remote monitoring)) AND (Mobile Health OR health OR mHealth OR eHealth OR Telehealth OR Telemedicine)) OR (Telehomecare)

2. (Scalability) OR (Feasibility) OR (Scaling up OR scale up OR upscale OR up-scale OR scale-up) OR ((Deployment OR Implementation OR Application) OR (Broad-scale OR Wide-scale OR Widespread OR Mainstream)) OR (((Efficienc*) AND (Program OR Intervention)) OR Economic* Viability)

3. (Follow-up Care* OR Follow Up Care* OR Care*) OR (Case Management OR Patient Care Planning)

4. ((Appraisal* OR Evaluation* OR Assessment* OR Appropriateness) AND ((Impact) OR (Cost-Effective* OR Qualitative OR Quantitative OR Index* OR Methodolog*) OR (Clinical Trial* AND (Pragmatic OR Naturalistic Randomized OR Practical OR Real World)) OR (Sustainability) OR (Profitability) OR (Risk*)))

5. #1 AND #2 AND #3 AND #4

Figure 1 illustrates the search performed in PubMed, Scopus, Web of Science, and EconLit covering studies from 2000 to 2020 (Figure 1 - Set #1). The authors chose to explore EconLit owing to the economic evaluation required to assess a health care intervention's scalability. The authors selected full-text and peer-reviewed papers written in English (Figure 1 - Set #2).

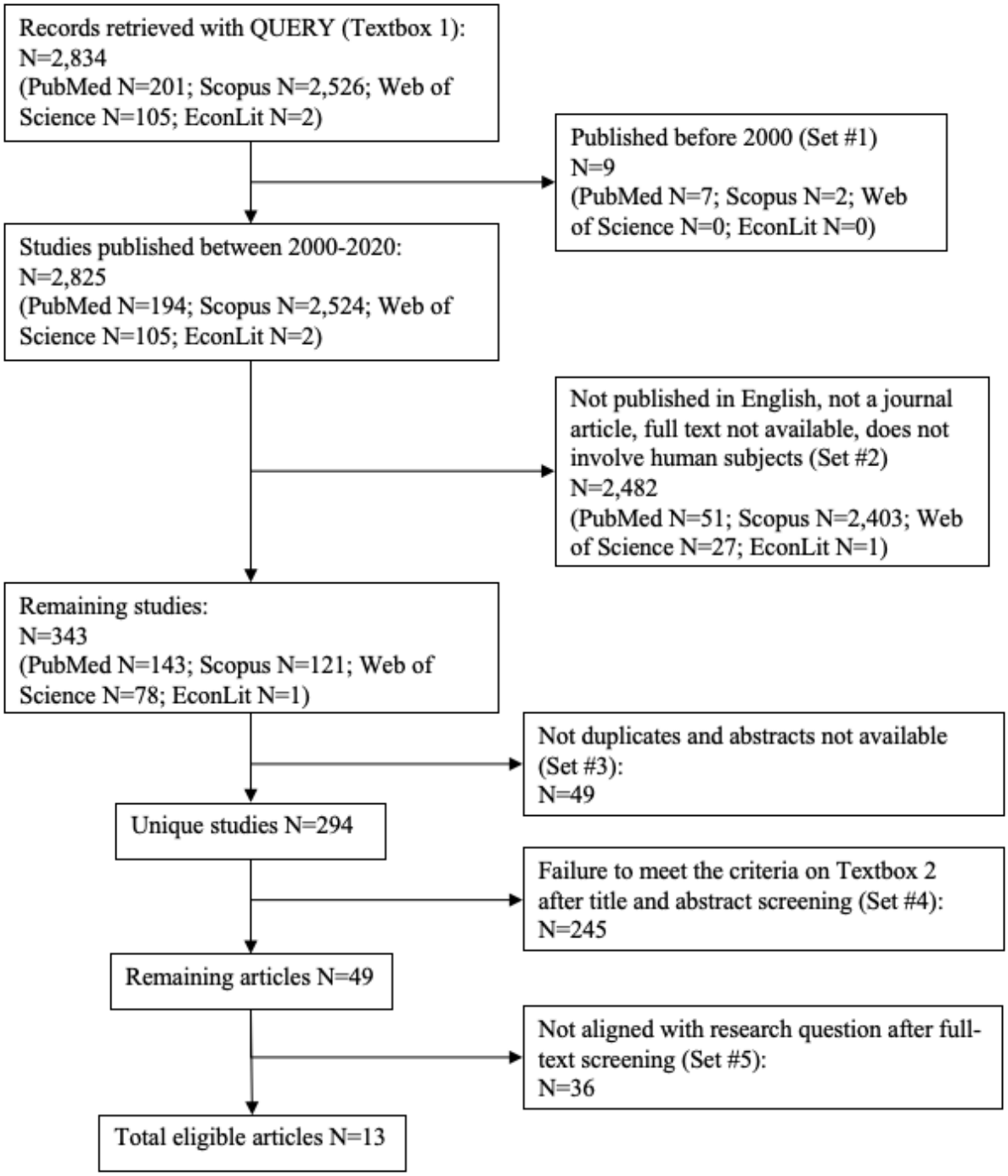
After removing the duplicates and references without abstracts (Figure 1 - Set #3), two authors independently scanned the titles and abstracts identified in the literature search and applied the selection criteria presented in Textbox 2 (Figure 1 - Set #4).

To guarantee that the article's topic aligned with the research question, the same authors scanned the 49 full-text articles, which reduced the number of studies considered for review to 13 (Set #5).

The authors analyzed 13 full-text articles, corresponding to 13 studies, in detail and registered all the observations in a literature matrix [23]. First, to better explore each intervention and summarize relevant, well-specified, and secure data, the authors conducted an integrative synthesis. The main variables were the country of origin, publication year, sample size, setting, duration of follow-up, comparator arms, type of technology, and study outcomes [24].

Second, the authors assessed the strength of each eligible study's evidence according to the 9-level classification system proposed by Jovell and Navarro-Rubio [25].

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram showing the included studies.



Textbox 2. Eligibility criteria for screening titles, abstracts, and full-text papers.

Inclusion criteria

- Health interventions shown to be efficacious on a small scale or under controlled conditions
- Assessment of the health intervention's ability to be expanded to real-world conditions to reach a more significant proportion of the eligible population while retaining effectiveness.
- Studies assessing at least one domain of scalability through the evaluation of feasibility, acceptability, costs, sustainability or, adaptability (ie, to suit the needs of the context in which it is to be scaled up)
- Described methods to assess the scalability of a health intervention

Exclusion criteria

- Telemonitoring involving invasive medical devices
- Studies that use telemonitoring "not involving the patients, their relatives, or informal caregivers, their relatives, or informal caregivers
- Studies that described the concept of scalability without providing an assessment method
- Studies just focusing on describing disease risk patterns or intervention efficacy testing
- Study protocols or medical testing procedures for potential scalability assessment and possible scale-up
- Statistical or conceptual modeling without a real-world study
- Facilitators and barriers to scale-up within specific interventions or general experiences of scale-up that did not provide a scalability assessment method
- Studies recommending an assessment method (of feasibility or acceptability or costs or sustainability or adaptability), but that did not assess the potential to scale-up a telemonitoring-based health intervention

Finally, they conducted narrative and thematic analyses to identify themes and patterns in the eligible articles and outline the findings under specific headings [24] to better examine how each study assessed the potential of scaling up an intervention. When disagreements occurred, the authors reached a consensus via discussion. One author extracted data from the studies and completed quotes, and the second author validated the data according to the definition of each category. The authors conducted this analysis based on the work undertaken by Milat and colleagues [18] in the development of a tool to perform systematic assessments of the suitability of health interventions for scale-up (ISAT). ISAT comprises three parts: setting the scene, planning the intervention implementation, and summarizing the scalability assessment. The first two parts made it possible to classify each study according to the stage of scale-up, context, and focus area. Moreover, Milat and colleagues' domains enabled the authors to identify the methods and instruments used by the researchers to assess the intervention's scalability [18].

The research conducted for each domain assessed in the eligible papers was classified as qualitative or quantitative. The research was classified as qualitative if it was based on the description of experiences, emotions, behaviors, events, or actions [26] and quantitative when the respective authors used numerical data to measure, categorize, or identify patterns, relationships, or generalizations through statistical analysis [26].

Results

Country of Origin and Year of Publication

From 2009 to 2020, the authors analyzed 13 studies in 7 countries, which focused on the potential to scale-up home telemonitoring health care interventions; however, more than half (n=7) were published between 2018 and 2020. Most of the articles (n=8) were from Canada and the United States, whereas the rest were from 5 European countries—Denmark (n=1), Italy (n=1), Lithuania (n=1), Netherlands (n=1), and Spain (n=1).

Population and Home Telemonitoring Intervention Assessment

Target Condition or Disease

The studies addressed either chronic or acute conditions, with a higher number of studies addressing only chronic conditions (n=8). The full spectrum of chronic conditions covered were cardiovascular diseases (n=4), chronic obstructive pulmonary diseases (n=2), cerebrovascular diseases (n=1), chronic obstructive sleep apnea (n=1), cystic fibrosis (n=1), and diabetes mellitus (gestational [n=1] and type 1 and 2 [n=1]). Further, one study only characterized the patients' condition as chronic or acute, and the remaining studies addressed multiple conditions (eg, surgical patients, cardiovascular and pulmonary diseases, diabetes mellitus).

Duration and Setting of Home Telemonitoring Intervention

Home telemonitoring was integrated into a follow-up service in the 13 studies and required a responsible health care professional (or a team) to manage the patient's care. The minimum duration of the follow-up was 3 consecutive nights (sleep apnea [27]). However, the 1-year (n=4) and 6-month (n=4) follow-up interventions were the most implemented. In particular, authors reporting the secondary prevention of cerebrovascular disease [28] defined the intervention according to recommended monitoring protocols, assuming a 20-year time horizon for the modeling strategy. Moreover, 10 studies had 2 dedicated teams for executing the intervention; one was

responsible for the patient's holistic care management and the other for telecare management. In two studies, the conventional care team was accountable for usual care and telecare management, and in the other, there was no traditional care team.

Types of Technologies

The technologies used in the studies ranged from a kit with just one regular telephone (1) to an integrated communication and data collection system with mobile devices (5). Moreover, six studies conducted home telemonitoring interventions with an integrated clinical data system, remote monitoring digital technology (mobile devices that collect physiological signs), and a telephone.

Study Design Assessment

Study Characteristics

The average total sample size of the studies was approximately 436 (maximum: 3086, minimum: 34), with an average treatment and control group size of 260.

To better understand the type of research conducted, it is essential to highlight that 6 out of the 13 studies were experimental. Therefore, the authors of these studies allocated participants to different treatment groups. As the other 7 studies were observational, there was no allocation of the participants. Most of the studies (n=10) were comparative studies (control group) with conventional care services, and the other 3 were single-arm studies.

Study Design Classification

According to the 9-level classification system proposed by Jovell and Navarro-Rubio [25], the studies conducted by Padwal and colleagues [28], and Vestergaard and colleagues [29] were classified as "very good," as they conducted randomized controlled trials with large samples. The studies by Lugo and colleagues [27], and Paré and colleagues [30] were classified as "good" as these studies were randomized controlled trials with small samples. Furthermore, the studies of Ware and colleagues [31], as well as Zaliūnas and colleagues [32], were classified as "poor" because they consisted of noncontrolled clinical series or descriptive studies. The other 7 were classified as fair and included nonrandomized controlled prospective studies (n=3), cohort studies (n=3), and case-control studies (n=1).

Scalability Assessment

Table 1 displays the scalability assessment domains for each study.

Table 1. Scalability assessment domains for each study.

Application field	Stage of scale-up	Domains for scale-up					Domains for implementation planning				
		Problem	Intervention	Context	Effectiveness	Costs and benefits	Fidelity and adaptability	Reach and acceptability	Setting and workforce	Infrastructure	Sustainability
Improved health outcomes in a rural area [33]	Pre–scale-up	Yes	Yes	Yes	Yes	Yes	No	Yes	No	No	No
Diabetes [34]	Pre–scale-up	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cystic fibrosis [35]	Pre–scale-up	Yes	Yes	No	Yes	Yes	No	No	Yes	Yes	Yes
Chronic heart failure [36]	Pre–scale-up	Yes	No	No	Yes	Yes	No	No	No	No	No
Obstructive sleep apnea [27]	Pre–scale-up	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Secondary prevention of cerebrovascular disease [28]	Pre–scale-up	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes
Heart failure [29]	Pre–scale-up	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No
Gestational diabetes mellitus [37]	Pre–scale-up	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Rural home health agencies [38]	Scale-up	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Chronic obstructive pulmonary disease [30]	Scale-up	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Ischemic heart disease [32]	Implementation	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Heart failure [31]	Implementation	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Chronic obstructive pulmonary disease [39]	Implementation	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes

Scale-Up Stages

The authors classified eight studies as being in the pre–scale-up stage because their descriptions consisted of steps or activities conducted before scaling up the evidence-based home telemonitoring intervention. Two studies described steps or actions involved in the dissemination of the intervention. The authors classified the other three studies as being in the implementation stage because their descriptions indicated using or integrating the evidence-based intervention within a setting.

Domains Considered for Scale-Up

Although all the studies described the problem under intervention and the target population, one study [36] did not provide details concerning the proposed home telemonitoring intervention to address the issue. All studies referred to the level of evidence available to support the proposed intervention's scale-up, either by referring to their work or other scientific literature., Three studies did not consider the known costs and benefits of delivering the intervention [31,32,39], and three more did not consider the strategic/political/environmental contexts that influence the scaling up of the intervention [27,35,36].

Domains Considered for Implementation Planning

Seven studies considered intervention changes when assessing fidelity, and nine studies assessed the level of acceptability perceived by the program deliverers or recipients of the intervention. Further, 9 studies referred to the definition of the intervention settings and the workforce required to scale-up, and 10 described the necessary infrastructure.

All the studies accounted for the sustainability of the home telemonitoring service by either referring to the long-term outcomes of the scale-up or the medium- and long-term sustainability of the intervention following scale-up.

Methods for Scalability Assessment

This section explains the research foci and methods used by the eligible studies in each domain of scalability assessment. When describing the problems, interventions, and contexts of their studies, all the researchers adopted qualitative research methods, as Table 2 shows. The definitions of the domains and research foci are given in Multimedia Appendix 1. We have included six publications [40-45] in this appendix.

Table 2. Qualitative studies on scalability assessment considering the problem, intervention, and context domains for scale-up.

Domain	Research focus	Research type	Data collection technique	Data analysis technique	Studies, n	Reference
Problem	Problem description	Qualitative	Document analysis	Narrative summary	13	[27-39,46-50]
Intervention	Intervention description	Qualitative	Document analysis	Narrative summary	12	[27-35,37-39]
Context	Context description	Qualitative	Document analysis	Narrative summary	10	[28-34,37-39]

All the studies adopted quantitative research methods to assess clinical outcomes namely surveys or questionnaires (n=10), published databases (n=2), and observations (n=1) (Table 3). To assess humanistic and satisfaction outcomes, the researchers chose surveys or questionnaires; however, for assessing for usage outcomes, they either conducted observations (n=9) or used published databases (n=3). As for validated instruments, only one was used in one study [27] to assess clinical outcomes, namely the Epworth Sleepiness Scale (ESS) [51]. For assessing humanistic outcomes, three validated questionnaires were used: EuroQol 5-Dimensions 5-Levels (EQ-5D-5L) [52] in the contexts of heart failure [29] and obstructive sleep apnea [27]; Quebec Sleep Questionnaire (QSQ) [53] for obstructive sleep apnea; and Chronic obstructive pulmonary disease Assessment Test (CAT) [54] for chronic obstructive pulmonary disease [39]. In the context of ischemic heart disease [32], two more validated questionnaires were used: Patient Satisfaction Questionnaire Form III (PSQIII) [55] and Thought Control Questionnaire (TCQ) [56].

Table 3. Quantitative research studies involving data analyses using descriptive and inferential statistics for scalability assessment considering the effectiveness domain for scale-up.

Research focus and data collection technique	Studies, n	Reference
Clinical outcome assessment		
Observation; published databases	3	[28,36,38]
Nonvalidated surveys or questionnaires	9	[29-35,37,39]
Validated surveys or questionnaires	1	[27]
Humanistic outcome assessment		
Nonvalidated surveys or questionnaires	3	[32-34]
Validated surveys or questionnaires	3	[27,29,39]
Satisfaction assessment		
Nonvalidated surveys or questionnaires	7	[27,29,30,33,34,37,39]
Validated surveys or questionnaires	1	[32]

For the domains of fidelity and acceptability, quantitative research methods involving observations were more predominantly used as the main data collection methods, as shown in Tables 4 and 5. Contrarily, for analyzing infrastructure, setting, and workforce, most of the studies chose qualitative techniques (n=8).

Table 4. Studies on scalability assessment concerning the reach and acceptability domain for implementation planning involving data analyses using descriptive and inferential statistics.

Research focus and type	Data collection technique	Studies, n	Reference
Acceptability assessment			
Quantitative	Observation	7	[27,30,32-34,37,39]
Qualitative	Semistructured interviews	1	[38]
Compliance assessment			
Quantitative	Nonvalidated surveys or questionnaires	1	[31]
Quantitative	Validated surveys or questionnaires	1	[32]
Penetration assessment			
Quantitative	Observation	2	[31,37]

Table 5. Research focus and methods found in the studies for scalability assessment concerning the fidelity and adaptability domain for implementation planning.

Research focus and type	Data collection technique	Data analysis technique	Studies, n	Reference
Adaptability assessment				
Quantitative	Observation	Descriptive statistics; inferential statistics	1	[27]
Qualitative	Observations; oral history or life stories	Narrative summary	2	[32,38]
Feasibility assessment				
Quantitative	Observation	Descriptive statistics; inferential statistics	2	[31,37]

When conducting economic evaluation (Table 6), the authors found 7 different types of techniques used across 10 studies (see Multimedia Appendix 2 for the main results of the studies that conducted economic evaluation of home telemonitoring). The most popular technique was cost-effectiveness analysis used in three studies with different fields of application. These three studies were able to show outcome improvements and cost savings. Table 7 presents the scalability assessment studies concerning the setting and workforce, infrastructure, and sustainability domains for implementation planning

Table 6. Quantitative research studies focusing on data collection using document screening and published databases for scalability assessment considering the costs and benefits domain for scale-up (research focus: economic evaluation).

Data analysis technique	Studies, n	Reference
Cost analysis	2	[34,35]
Cost-benefit	1	[38]
Cost-effectiveness	3	[27,33,37]
Cost minimization	1	[30]
Cost utility	2	[28,29]
Cost-saving simulation	1	[35]
Value of information analysis	1	[36]

Table 7. Studies on scalability assessment concerning the setting and workforce, infrastructure, and sustainability domains for implementation planning.

Domain and research focus and type	Data collection technique	Data analysis technique	Studies, n	Reference
Setting and workforce				
Setting and workforce assessment				
Qualitative	Observations; oral history or life stories	Narrative summary	8	[27,28,31,32,34,35,37,39]
Quantitative	Observation	Descriptive statistics	1	[30]
Infrastructure				
Infrastructure assessment				
Qualitative	Observations; oral history or life stories	Narrative summary	9	[27,28,30,32,34,35,37-39]
Qualitative	Semistructured Interviews	Descriptive statistics	1	[31]
Sustainability				
Opportunity and challenge assessment				
Qualitative	Observations; oral history or life stories	Narrative summary	12	[27-30,32-39]
Qualitative	Semistructured interviews	Narrative summary	1	[31]

Scalability Assessment

All the 13 articles assessed scalability based on the results achieved in the respective studies. Table 8 summarizes the assessments obtained through narrative analysis. On the one hand, two studies provided positive assessments regarding the potential to scale-up the intervention. On the other hand, eight studies highlighted the need for cost-effectiveness or cost-benefit analysis before proceeding to scale-up the intervention.

Table 8. Scalability assessment based on the authors' conclusions in each study.

Scalability assessment	Studies, N	Reference
Not able to be expanded	1	[27]
Able to be expanded, *but* the diffusion and sustainability will depend on a supportive policy environment	1	[34]
Able to be expanded *but* requires cost-benefit analysis for reimbursement planning	3	[28,36,38,39]
Able to be expanded *but* requires cost-effectiveness analysis	3	[29,30,35]
Able to be expanded *but* requires some technical changes, cost-benefit analysis for reimbursement planning, and solutions for regulatory issues	2	[32,33]
Able to be expanded under real-world conditions	2	[31,37]

Discussion

Principal Results

Despite the rapid growth of telemedicine applications in the last few years, particularly after the emergence of COVID-19, scientific studies assessing the scalability of these health interventions are scarce [19].

In this review, all the eligible studies are from developed countries, particularly the United States and Canada. The absence of such studies in developing countries could be owing to the lack of specialized human resources, information and communications technology (ICT) infrastructure, and equipment [46]. Besides, the significant difference found between North America and Europe might be related to the requirement of evidence to justify private payer reimbursement for health care interventions [47] or the investment in developing strategies to encourage telemedicine adoption [48]. Nevertheless, this review has not identified studies from countries that invested significantly in telehealth solutions, such as the United Kingdom or Australia [46]. The justification for this might be the frequent research focus of health interventions on clinical effectiveness [11], instead of assessing their scale-up potential. More than half of the studies were published between 2018 and 2020. Thus, this research area is receiving more attention from the scientific community as a logical next step after demonstrating robust evidence regarding the effectiveness and technological maturity of such interventions.

The use of one of the most recent scalability assessment frameworks [18] granted the opportunity to compare the strategies used to assess the scale-up potentials of interventions in each study. This advantage of this framework is that it allows the analysis of different domains considering the stage of the

transference process of an intervention from a research setting into the practical implementation stage.

This review suggests an agreement in some analyzed domains, such as problems, interventions, effectiveness, costs, and benefits, to support the decision to scale-up interventions. However, this is not the case for the methods and instruments used. For example, although cost-effectiveness was the most common approach across the 13 studies, researchers used 7 different cost analysis methods. Moreover, to demonstrate effectiveness, studies provided evidence of different outcomes, such as clinical, humanistic, and utilization outcomes. This inconsistency leads to different scalability assessments and does not enable comparing interventions with home telemonitoring technologies.

There is a recognized methodological gap in understanding other relevant domains such as the sociopolitical context, setting, workforce, and implementation infrastructure to provide the home telemonitoring intervention to the target population. A common framework will allow determining if interventions demonstrated as effective are appropriate and feasible in other settings [18,49].

Lastly, another relevant result obtained from this systematic review was that researchers assigned different weights to the analyzed domains when concluding the intervention scalability. On the one hand, 12 studies concluded their ability to scale- up based on the costs and outcomes of the interventions, although they had analyzed other domains. On the other hand, one study restrained the decision to scale-up the intervention based on the policy environment. Future research should address the influence that each domain has on the final decision to scale-up the interventions with sound and transparent methods, avoiding mistakes reported in the literature [50].

Limitations

This relevant limitation of this review might be associated with the low maturity of this research area, despite its recent growth. Additionally, one database filter concerned peer-reviewed journals, which influenced the rejection of studies with no statistical significance but could have been relevant in this review with respect to the domains and methods used when assessing scalability. This review only considered studies published in English, which might have influenced the number of eligible studies. Moreover, the authors did not conduct a meta-analysis owing to the limited number of studies on this subject. Finally, the domains used to analyze the scalability assessment strategies were predefined, thus limiting the spectrum of domains studied.

Conclusions

Studies on home telemonitoring interventions integrated into follow-up care have already proved their efficacy. Although some studies focused on including domains such as effectiveness, costs, and benefits, these are not enough to assess the potential of scaling up these interventions. As technology progresses and the need for providing care to more people in their homes increases, it is extremely important to conduct more studies on scalability assessment considering domains such as workforce and infrastructure characteristics and the strategic context. Future research should establish rigorous study designs and scientific methods to assess scalability based on the results of this systematic review. Further understanding of the usage of health services and medium- and long-term sustainability of interventions would yield more robust evidence to support their future integration into mainstream care delivery systems. This research area, although still emerging, will advance knowledge on the factors that influence the successful scale-up of interventions.

Acknowledgments

We are thankful to Professor Mónica Oliveira for her support and guidance in the ideation process of this systematic review. In addition, we are grateful for the funding provided by Portugal 2020, under the scope of PROJETOS DE I&DT EMPRESAS EM COPROMOÇÃO - COVID-19 (15/SI/2020) and the project Fast Track Covid, LISBOA-01-02B7-FEDER-069202.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Glossary of research methods and scalability assessment domains used to systematically review the eligible studies in this work.
[PDF File (Adobe PDF File), 82 KB - mhealth_v9i8e29381_app1.pdf]

Multimedia Appendix 2
Main results of the economic evaluations conducted in each eligible study that addressed the domains of costs and benefits.
[PDF File (Adobe PDF File), 77 KB - mhealth_v9i8e29381_app2.pdf]

References

1. Annis T, Pleasants S, Hultman G, Lindemann E, Thompson J, Billecke S, et al. Rapid implementation of a COVID-19 remote patient monitoring program. J Am Med Inform Assoc 2020 Aug;27(8):1326-1330 [FREE Full text] [doi: 10.1093/jamia/ocaa097] [Medline: 32392280]

2. Adolph C, Amano K, Bang-Jensen B, Fullman N, Wilkerson J. Pandemic politics: timing state-level social distancing responses to COVID-19. J Health Polit Policy Law 2021 Apr;46(2):211-233. [doi: 10.1215/03616878-8802162] [Medline: 32955556]
3. Anderson M, Mckee M, Mossialos E. Developing a sustainable exit strategy for COVID-19: health, economic and public policy implications. J R Soc Med 2020 May;113(5):176-178. [doi: 10.1177/0141076820925229] [Medline: 32407649]
4. European Centre for Disease Prevention and Control (ECDC) 2021. URL: https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases [accessed 2021-08-02]
5. Carter P, Anderson M, Mossialos E. Health system, public health, and economic implications of managing COVID-19 from a cardiovascular perspective. Eur Heart J 2020 Jul;41(27):2516-2518 [FREE Full text] [doi: 10.1093/eurheartj/ehaa342] [Medline: 32320040]
6. Srivatana V, Liu F, Levine DM, Kalloo SD. Early use of telehealth in home dialysis during the COVID-19 pandemic in New York City. Kidney360 2020 Apr;1(6):524-526. [doi: 10.34067/kid.0001662020]
7. Fisk M, Livingstone A, Pit SW. Telehealth in the context of COVID-19: changing perspectives in Australia, the United Kingdom, and the United States. J Med Internet Res 2020 Jun;22(6):e19264 [FREE Full text] [doi: 10.2196/19264] [Medline: 32463377]
8. Smith AC, Thomas E, Snoswell CL, Haydon H, Mehrotra A, Clemensen J, et al. Telehealth for global emergencies: implications for coronavirus disease 2019 (COVID-19). J Telemed Telecare 2020 Mar;26(5):309-313. [doi: 10.1177/1357633x20916567]
9. Strode SW, Gustke S, Allen A. Technical and clinical progress in telemedicine. JAMA 1999;281(12):1066-1068. [doi: 10.1001/jama.281.12.1066] [Medline: 10188642]
10. Meystre S. The current state of telemonitoring: a comment on the literature. Telemed J E Health 2005 Feb;11(1):63-69. [doi: 10.1089/tmj.2005.11.63] [Medline: 15785222]
11. Pare G, Jaana M, Sicotte C. Systematic review of home telemonitoring for chronic diseases: the evidence base. J Am Med Inform Assoc 2007 May;14(3):269-277. [doi: 10.1197/jamia.m2270]
12. Mengden T, Vetter H, Tislér A, Illyés M. Tele-monitoring of home blood pressure. Blood Press Monit 2001 Aug;6(4):185-189. [doi: 10.1097/00126097-200108000-00004] [Medline: 11805466]
13. Riley JP, Gabe JP, Cowie MR. Does telemonitoring in heart failure empower patients for self-care? a qualitative study. J Clin Nurs 2013 Sep;22(17-18):2444-2455. [doi: 10.1111/j.1365-2702.2012.04294.x] [Medline: 23185992]
14. Makhni S, Atreja A, Sheon A, Van Winkle B, Sharp J, Carpenter N. The broken health information technology innovation pipeline: a perspective from the NODE health consortium. Digit Biomark 2017 Aug;1(1):64-72 [FREE Full text] [doi: 10.1159/000479017] [Medline: 32095746]
15. Milat AJ, King L, Bauman AE, Redman S. The concept of scalability: increasing the scale and potential adoption of health promotion interventions into policy and practice. Health Promot Int 2013 Sep;28(3):285-298. [doi: 10.1093/heapro/dar097] [Medline: 22241853]
16. European C. Defining value in "value-based healthcare": report of the expert panel on effective ways of investing in health (EXPH). Publications Office of the European Union. 2019 Jul. URL: https://op.europa.eu/s/psov [accessed 2021-08-02]
17. Haynes B. Can it work? Does it work? Is it worth it? the testing of healthcare interventions is evolving. BMJ 1999 Sep;319(7211):652-653 [FREE Full text] [doi: 10.1136/bmj.319.7211.652] [Medline: 10480802]
18. Milat A, Lee K, Conte K, Grunseit A, Wolfenden L, van Nassau F, et al. Intervention scalability assessment tool: a decision support tool for health policy makers and implementers. Health Res Policy Sys 2020 Jan;18(1):1 [FREE Full text] [doi: 10.1186/s12961-019-0494-2] [Medline: 31900230]
19. Kidholm K, Ekeland AG, Jensen LK, Rasmussen J, Pedersen CD, Bowes A, et al. A model for assessment of telemedicine applications: mast. Int J Technol Assess Health Care 2012 Jan;28(1):44-51. [doi: 10.1017/S0266462311000638] [Medline: 22617736]
20. Ben Charif A, Zomahoun HTV, LeBlanc A, Langlois L, Wolfenden L, Yoong SL, et al. Effective strategies for scaling up evidence-based practices in primary care: a systematic review. Implement Sci 2017 Nov;12(1):139 [FREE Full text] [doi: 10.1186/s13012-017-0672-y] [Medline: 29166911]
21. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
22. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions. 2nd Edition. Chichester, United Kindgom: John Wiley & Sons; 2019.
23. Kruse CS. Writing a systematic review for publication in a health-related degree program. JMIR Res Protoc 2019 Oct;8(10):e15490 [FREE Full text] [doi: 10.2196/15490] [Medline: 31527018]
24. Dixon-Woods M, Agarwal S, Jones D, Young B, Sutton A. Synthesising qualitative and quantitative evidence: a review of possible methods. J Health Serv Res Policy 2005 Jan;10(1):45-53. [doi: 10.1177/135581960501000110] [Medline: 15667704]
25. Jovell AJ, Navarro-Rubio MD. Evaluation of scientific evidence. Med Clin (Barc) 1995 Dec;105(19):740-743. [Medline: 8523956]

26. University of Newcastle Library Guides - Research Methods. University of Newcastle Library Guides. URL: https://libguides.newcastle.edu.au/researchmethods [accessed 2021-08-02]
27. Lugo VM, Garmendia O, Suarez-Girón M, Torres M, Vázquez-Polo FJ, Negrín MA, et al. Comprehensive management of obstructive sleep apnea by telemedicine: clinical improvement and cost-effectiveness of a virtual sleep unit. a randomized controlled trial. PLoS One 2019;14(10):e0224069 [FREE Full text] [doi: 10.1371/journal.pone.0224069] [Medline: 31647838]
28. Padwal RS, So H, Wood PW, Mcalister FA, Siddiqui M, Norris CM, et al. Cost-effectiveness of home blood pressure telemonitoring and case management in the secondary prevention of cerebrovascular disease in Canada. J Clin Hypertens (Greenwich) 2019 Feb;21(2):159-168 [FREE Full text] [doi: 10.1111/jch.13459] [Medline: 30570200]
29. Vestergaard AS, Hansen L, Sørensen SS, Jensen MB, Ehlers LH. Is telehealthcare for heart failure patients cost-effective? an economic evaluation alongside the Danish TeleCare North heart failure trial. BMJ Open 2020 Jan;10(1):e031670 [FREE Full text] [doi: 10.1136/bmjopen-2019-031670] [Medline: 31992604]
30. Paré G, Poba-Nzaou P, Sicotte C, Beaupré A, Lefrançois É, Nault D, et al. Comparing the costs of home telemonitoring and usual care of chronic obstructive pulmonary disease patients: a randomized controlled trial. European Research in Telemedicine / La Recherche Européenne en Télémédecine 2013 Jun;2(2):35-47. [doi: 10.1016/j.eurtel.2013.05.001]
31. Ware P, Ross HJ, Cafazzo JA, Laporte A, Gordon K, Seto E. Evaluating the implementation of a mobile phone-based telemonitoring program: longitudinal study guided by the consolidated framework for implementation research. JMIR Mhealth Uhealth 2018 Jul;6(7):e10768 [FREE Full text] [doi: 10.2196/10768] [Medline: 30064970]
32. Zaliūnas R, Benetis R, Vanagas G, Slapikas R, Vainoras A. Implementation of international transtelephonic ECG platform for patients with ischemic heart disease. Medicina (Kaunas) 2009;45(2):104-110 [FREE Full text] [Medline: 19289900]
33. Hicks LL, Fleming DA, Desaulnier A. The application of remote monitoring to improve health outcomes to a rural area. Telemed J E Health 2009 Sep;15(7):664-671. [doi: 10.1089/tmj.2009.0009] [Medline: 19694598]
34. Nundy S, Dick JJ, Chou C, Nocon RS, Chin MH, Peek ME. Mobile phone diabetes project led to improved glycemic control and net savings for Chicago plan participants. Health Aff (Millwood) 2014 Feb;33(2):265-272 [FREE Full text] [doi: 10.1377/hlthaff.2013.0589] [Medline: 24493770]
35. Tagliente I, Trieste L, Solvoll T, Murgia F, Bella S. Telemonitoring in cystic fibrosis: a 4-year assessment and simulation for the next 6 years. Interact J Med Res 2016 May;5(2):e11 [FREE Full text] [doi: 10.2196/ijmr.5196] [Medline: 27142963]
36. Grustam AS, Buyukkaramikli N, Koymans R, Vrijhoef HJM, Severens JL. Value of information analysis in telehealth for chronic heart failure management. PLoS One 2019;14(6):e0218083 [FREE Full text] [doi: 10.1371/journal.pone.0218083] [Medline: 31220101]
37. Lemelin A, Paré G, Bernard S, Godbout A. Demonstrated cost-fffectiveness of a telehomecare program for gestational diabetes mellitus management. Diabetes Technol Ther 2020 Mar;22(3):195-202. [doi: 10.1089/dia.2019.0259] [Medline: 31603351]
38. Hansen D, Golbeck AL, Noblitt V, Pinsonneault J, Christner J. Cost factors in implementing telemonitoring programs in rural home health agencies. Home Healthc Nurse 2011 Jun;29(6):375-382. [doi: 10.1097/NHH.0b013e31821b736f] [Medline: 21633229]
39. Farias R, Sedeno M, Beaucage D, Drouin I, Ouellet I, Joubert A, et al. Innovating the treatment of COPD exacerbations: a phone interactive telesystem to increase COPD action plan adherence. BMJ Open Respir Res 2019 Apr;6(1):e000379 [FREE Full text] [doi: 10.1136/bmjresp-2018-000379] [Medline: 31178998]
40. Walton MK, Powers JH, Hobart J, Patrick D, Marquis P, Vamvakas S, et al. Clinical outcome assessments: conceptual foundation—report of the ISPOR clinical outcomes assessment – emerging good practices for Outcomes Research Task Force. Value Health 2015 Sep;18(6):741-752 [FREE Full text] [doi: 10.1016/j.jval.2015.08.006] [Medline: 26409600]
41. Velentgas P, Dreyer NA, Nourjah P, Smith SR, Torchia MM, editors. Chapter 6-Outcome definition and measurement. In: Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide. Rockville, United States: Agency for Healthcare Research and Quality; Jan 2013.
42. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. Fourth edition. Oxford: Oxford University Press; 2015.
43. Moncher FJ, Prinz RJ. Treatment fidelity in outcome studies. Clin Psychol Rev 1991 Jan;11(3):247-266. [doi: 10.1016/0272-7358(91)90103-2]
44. Eldridge SM, Lancaster GA, Campbell MJ, Thabane L, Hopewell S, Coleman CL, et al. Defining feasibility and pilot studies in preparation for randomised controlled trials: development of a conceptual framework. PLoS One 2016;11(3):e0150205 [FREE Full text] [doi: 10.1371/journal.pone.0150205] [Medline: 26978655]
45. Haynes RB, Taylor DW, Sackett DL, editors. Compliance in Health Care. Baltimore: Johns Hopkins University Press; 1979.
46. Combi C, Pozzani G, Pozzi G. Telemedicine for developing countries. a survey and some design issues. Appl Clin Inform 2016 Nov;7(4):1025-1050 [FREE Full text] [doi: 10.4338/ACI-2016-06-R-0089] [Medline: 27803948]
47. Whitten P, Buis L. Private payer reimbursement for telemedicine services in the United States. Telemed J E Health 2007 Feb;13(1):15-23. [doi: 10.1089/tmj.2006.0028] [Medline: 17309350]
48. Gheorghiu B, Ratchford F. Scaling up the use of remote patient monitoring in Canada. Stud Health Technol Inform 2015;209:23-26. [Medline: 25980701]

49. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
50. Keeney RL. Value-focused thinking: Identifying decision opportunities and creating alternatives. Eur J Oper Res 1996 Aug;92(3):537-549. [doi: 10.1016/0377-2217(96)00004-5]
51. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep 1991 Dec;14(6):540-545. [doi: 10.1093/sleep/14.6.540] [Medline: 1798888]
52. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011 Dec;20(10):1727-1736 [FREE Full text] [doi: 10.1007/s11136-011-9903-x] [Medline: 21479777]
53. Lacasse Y, Bureau M, Sériès F. A new standardised and self-administered quality of life questionnaire specific to obstructive sleep apnoea. Thorax 2004 Jun;59(6):494-499 [FREE Full text] [doi: 10.1136/thx.2003.011205] [Medline: 15170032]
54. Jones PW, Harding G, Berry P, Wiklund I, Chen W, Kline Leidy N. Development and first validation of the COPD assessment test. Eur Respir J 2009 Sep;34(3):648-654 [FREE Full text] [doi: 10.1183/09031936.00102509] [Medline: 19720809]
55. Ware JE, Snyder MK, Wright W, Davies AR. Defining and measuring patient satisfaction with medical care. Eval Program Plann 1983 Jan;6(3-4):247-263. [doi: 10.1016/0149-7189(83)90005-8] [Medline: 10267253]
56. Wells A, Davies MI. The thought control questionnaire: a measure of individual differences in the control of unwanted thoughts. Behav Res Ther 1994 Nov;32(8):871-878. [doi: 10.1016/0005-7967(94)90168-6] [Medline: 7993332]

Abbreviations

CAT: Chronic Obstructive Pulmonary Disease Assessment Test
EQ-5D-5L: EuroQol 5-Dimensions 5-Levels
ESS: Epworth Sleepiness Scale
ICT: information and communications technology
ISAT: Intervention Scalability Assessment Tool
MeSH: Medical Subject Heading
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis
PSQIII: Patient Satisfaction Questionnaire Form III
QSQ: Quebec Sleep Questionnaire
TCQ: Thought Control Questionnaire

Edited by L Buis; submitted 05.04.21; peer-reviewed by L Trieste, S Pit; comments to author 11.06.21; revised version received 25.06.21; accepted 14.07.21; published 19.08.21.

Please cite as:
Azevedo S, Rodrigues TC, Londral AR
Domains and Methods Used to Assess Home Telemonitoring Scalability: Systematic Review
JMIR Mhealth Uhealth 2021;9(8):e29381
URL: https://mhealth.jmir.org/2021/8/e29381
doi:10.2196/29381
PMID:34420917

Original Paper

Factors Associated With Intention to Adopt mHealth Apps Among Dementia Caregivers With a Chronic Condition: Cross-sectional, Correlational Study

Kyra Jennifer Waligora Mendez[1], BSN, PhD; Chakra Budhathoki[1], PhD; Alain Bernard Labrique[2], PhD; Tatiana Sadak[3], PhD, FAAN; Elizabeth K Tanner[1], PhD, FAAN; Hae Ra Han[1,2], PhD, FAAN

[1]School of Nursing, Johns Hopkins University, Baltimore, MD, United States

[2]Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States

[3]School of Nursing, University of Washington, Seattle, WA, United States

Corresponding Author:
Kyra Jennifer Waligora Mendez, BSN, PhD
School of Nursing
Johns Hopkins University
525 N Wolfe St
Baltimore, MD, 21205
United States
Phone: 1 410 929 4327
Email: kwaligo1@jhu.edu

Abstract

Background: In the United States, nearly 80% of family caregivers of people with dementia have at least one chronic condition. Dementia caregivers experience high stress and burden that adversely affect their health and self-management. mHealth apps can improve health and self-management among dementia caregivers with a chronic condition. However, mHealth app adoption by dementia caregivers is low, and reasons for this are not well understood.

Objective: The purpose of this study is to explore factors associated with dementia caregivers' intention to adopt mHealth apps for chronic disease self-management.

Methods: We conducted a cross-sectional, correlational study and recruited a convenience sample of dementia caregivers. We created a survey using validated instruments and collected data through computer-assisted telephone interviews and web-based surveys. Before the COVID-19 pandemic, we recruited dementia caregivers through community-based strategies, such as attending community events. After nationwide closures due to the pandemic, the team focused on web-based recruitment. Multiple logistic regression analyses were used to test the relationships between the independent and dependent variables.

Results: Our sample of 117 caregivers had an average age of 53 (SD 17.4) years, 16 (SD 3.3) years of education, and 4 (SD 2.5) chronic conditions. The caregivers were predominantly women (92/117, 78.6%) and minorities (63/117, 53.8%), experienced some to extreme income difficulties (64/117, 54.7%), and were the child or child-in-law (53/117, 45.3%) of the person with dementia. In logistic regression models adjusting for the control variables, caregiver burden (odds ratio [OR] 1.3, 95% CI 0.57-2.8; *P*=.57), time spent caregiving per week (OR 1.7, 95% CI 0.77-3.9; *P*=.18), and burden of chronic disease and treatment (OR 2.3, 95% CI 0.91-5.7; *P*=.08) were not significantly associated with the intention to adopt mHealth apps. In the final multiple logistic regression model, only perceived usefulness (OR 23, 95% CI 5.6-97; *P*<.001) and the interaction term for caregivers' education and burden of chronic disease and treatment (OR 31, 95% CI 2.2-430; *P*=.01) were significantly associated with their intention to adopt mHealth apps. Perceived ease of use (OR 2.4, 95% CI 0.67-8.7; *P*=.18) and social influence (OR 1.8, 95% CI 0.58-5.7; *P*=.31) were not significantly associated with the intention to adopt mHealth apps.

Conclusions: When designing mHealth app interventions for dementia caregivers with a chronic condition, it is important to consider caregivers' perceptions about how well mHealth apps can help their self-management and which app features would be most useful for self-management. Caregiving factors may not be relevant to caregivers' intention to adopt mHealth apps. This is promising because mHealth strategies may overcome barriers to caregivers' self-management. Future research should investigate reasons why caregivers with a low education level and low burden of chronic disease and treatment have significantly lower intention to adopt mHealth apps for self-management.

(JMIR Mhealth Uhealth 2021;9(8):e27926) doi:10.2196/27926

KEYWORDS

mHealth applications; mobile health; intention to adopt mHealth applications; dementia caregivers; family caregiving; chronic disease self-management; mobile phone

Introduction

Background

In the United States, more than 11 million family caregivers provide care to a loved one with Alzheimer disease or related dementias [1,2]. Up to 80% of the caregivers have chronic health conditions [3,4]. However, because of the high demands of caregiving responsibilities, caregivers experience challenges with their own self-management [5]. Self-management is an individual's ability to manage or cope with the physical, psychosocial, and cultural effects of living with a chronic health condition [6].

Previous research supports that family caregivers of people with dementia perform less self-management than noncaregivers and experience worse health and well-being outcomes [7-10]. High caregiver burden and stress are barriers to self-management for family caregivers of people with dementia [5,11,12]. The COVID-19 pandemic has further exacerbated challenges to caregivers' self-management, with preliminary research reporting that the pandemic has increased anxiety and strain among family caregivers [13,14]. In addition, family caregivers are experiencing poorer mental and physical health outcomes than noncaregivers during the COVID-19 pandemic [15]. This further highlights the critical need for innovative methods that are readily accessible and improve caregiver self-management, health, and well-being.

Literature Review

mHealth strategies are effective in improving self-management and health outcomes of persons living with diabetes, mental health conditions, and cancer, among other chronic conditions [16-19]. However, family caregivers are less likely to use mobile apps for health-related needs than the general population [20], and fewer than 50% of the dementia caregivers use mHealth apps for their own health [21]. The reasons for these findings are largely unknown and require additional study [20,21].

The Technology Acceptance Model (TAM) is a well-known theoretical framework for exploring the factors associated with mHealth app adoption. The TAM was originally developed to explain the intention to adopt software systems [22] but has since been adapted to explore mHealth app adoption [23,24]. The TAM posits 2 technological factors that predict intention to adopt technology are perceived usefulness (beliefs about how well mHealth apps will help oneself to perform self-management) and perceived ease of using technology (one's beliefs that using mHealth apps will take little effort) [24-26]. Perceived usefulness and perceived ease of use have been positively associated with the intention to adopt mHealth solutions among persons with a chronic condition [23,24,27] and with dementia caregivers' intention to adopt wearable devices to manage persons with dementia [28] and caregiving-supportive technologies [29].

Prior studies have expanded the TAM to improve its utility and predictive power [23,24,28,30-32]. For example, in noncaregiver populations living with a chronic condition, social influence (perceptions that people who are important in your life believe that you should use technology) and perceptions of chronic disease threats have been associated with the intention to adopt mHealth solutions such as apps [23,24,30,31]. Perceptions that caregiving mHealth apps can prevent threats to the care recipient's health have also been associated with the intention to adopt mHealth apps among caregivers [32]. Nevertheless, it is unclear how caregivers' own chronic disease threats or burden may influence their intention to adopt self-management mHealth apps.

Furthermore, the findings from other studies suggest that caregiving factors may be relevant to caregivers' intention to adopt mHealth apps. For example, in the context of mHealth apps that support caregiving, caregivers with higher caregiver burden and strain had higher mHealth app use [33], and mHealth app use reduced caregiver strain and depression [34]. However, it is unclear if these caregiving factors are relevant to caregivers' use of mHealth apps for self-management. As caregiver burden and hours providing care per week are barriers to caregivers' self-management [5,12], it is important to further explore how these caregiving factors may affect caregivers' use of mHealth apps for their self-management.

In addition, racial and ethnic groups have similar rates of smartphone ownership according to national surveys [35,36], with Hispanic and Asian households having slightly higher smartphone ownership [36]; however, there are differences in whether they have downloaded an mHealth app [37,38]. Other studies have supported the existence of income and education differences in mobile device use [37,38], but that education may be a more comprehensive predictor of electronic health use than income [38]. Thus, it is also important to explore how the factors associated with mHealth adoption may differ by race or ethnicity and education to address disparities in mHealth app adoption.

Objectives

Taken together, although much progress has been made in expanding the TAM, there is still limited knowledge of the factors associated with mHealth app adoption among dementia caregivers with a chronic condition. Caregivers are often burdened to care for their own chronic health conditions, in addition to the multimorbidities of the person with dementia, and therefore have unique barriers to self-management compared with other populations [5]. To our knowledge, there are no prior studies that have investigated the factors associated with the intention to adopt mHealth apps for self-management among caregivers living with a chronic health condition. To fill this gap, the purpose of this study is to understand factors related to the intention of family caregivers of people with dementia

to adopt mHealth apps for their own chronic disease self-management. The study aims are as follows:

- Aim 1: to examine the relationships among dementia caregivers' technological, self-management, and caregiving factors and their intention to adopt mHealth apps for self-management. Hypothesis 1: we hypothesized that technological and self-management factors would be positively, and caregiving factors would be negatively, associated with the intention to adopt mHealth apps for chronic disease self-management, controlling for the caregivers' multimorbidities, age, gender, and income.
- Aim 2: to explore whether the caregivers' race or ethnicity and education moderate the relationship between the study variables and caregivers' intention to adopt mHealth apps for chronic disease self-management.

Methods

Study Design and Sample

We conducted a cross-sectional, correlational study and collected data in English and Spanish using computer-assisted telephone interviews and a web-based survey, both of which used the same web-based REDCap (Research Electronic Data Capture [39]) survey. Individuals were eligible for the study if they met the following criteria: aged 18 years or older; caring for a family member or friend with Alzheimer disease or related dementias; living with a chronic health condition; able to speak and understand English or Spanish; and owns, or has access to, a mobile device. Family caregivers were excluded if they, or the persons with dementia being given care, were institutionalized.

Using G*Power version 3.1.9.2 (Heinrich Heine University) and effect sizes from a recent study [23], we estimated that a sample size of 110 was needed for 85% power to detect a medium effect size with α=.05 for 2-sided tests. We also aimed to oversample minority caregivers by stratifying study recruitment. We doubled the population-based proportions of each racial or ethnic group [1] and planned to recruit 30 Black or African American, 25 Hispanic or Latino, and 11 Asian caregivers.

Procedures

All study procedures were approved by the Johns Hopkins Medicine Institutional Review Board (IRB). The study survey was created and piloted with content experts. After entering it into REDCap, it was piloted on the web and over the phone with community members to ensure that the skip patterns, survey flow, and instructions were appropriate before implementation. As part of the survey, the team provided pictures of an evidence-based self-management mHealth app for persons with diabetes to standardize the caregivers' conception of an mHealth app [40,41].

Data were collected in English from June 2019 to August 2020 and in Spanish from July 2020 to August 2020 (see Multimedia Appendix 1 for CHERRIES [Checklist for Reporting Results of Internet E-Surveys] checklist [42]). We recruited a convenience sample using community- and web-based methods [43]. Before the COVID-19 pandemic, the recruitment efforts focused on the Baltimore-Washington metropolitan area. After the nationwide lockdowns in March 2020, the team concentrated on web-based recruitment strategies. The study team members contacted local support groups, attended community-based events, received referrals from an Alzheimer disease treatment center and research center, and placed local newspaper advertisements. We also registered the study on the web with the Alzheimer's Association TrialMatch service and the Clinical Trials Finder of the National Institute on Aging. These methods required people to contact the study team, be referred, or sign up to be contacted to participate. After receiving the referrals or contacts of interested people, a team member screened them for eligibility and completed a phone interview or sent a personalized link to the web-based survey, which could only be completed once.

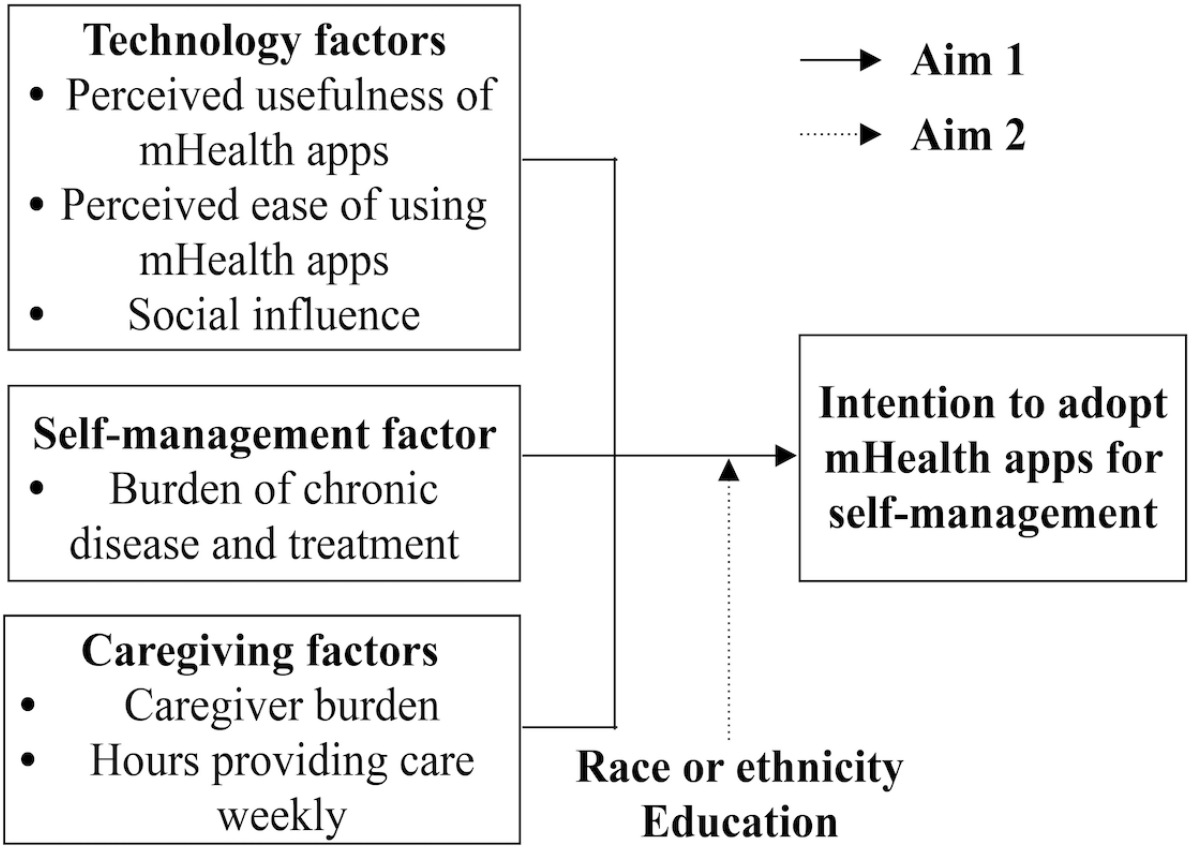
In addition, the team recruited on the web by posting advertisements on a Johns Hopkins University online news center and on social media (Google, Facebook, and YouTube) and by sending recruitment emails through a web-based research registry (ResearchMatch). These methods included an anonymous link to the eligibility screening survey. Interested individuals could click the link, complete the eligibility survey, and begin the web-based survey, if eligible. All eligible participants received information on the study purpose, procedures, risks, and benefits and consented to participate through IRB-approved oral or web-based consent. Data were stored in the REDCap database, to which only authorized, IRB-approved team members with password-protected accounts had access. All participants who completed the study survey were remunerated with a US $10 gift card.

Study Variables and Instruments

The theoretical framework guiding the study was an expanded TAM, which included the factors relevant to caregivers and their self-management [5,24,33,34]. The theoretical framework included the technology-related factors from the original and expanded TAM, caregiving factors, and a self-management factor (burden of chronic disease and treatment, defined as how much the caregivers' chronic condition and its treatment impacts daily life). It also proposed that education and race or ethnicity moderate the relationships between technology, self-management, and caregiving factors and the intention to adopt mHealth apps (Figure 1).

To measure the sociodemographic variables, we used questions from the US Census and national surveys. Income was captured with a well-validated question of financial strain ("How hard is it for you to pay for the very basics like food, housing, medical care, and heating?") [44]. Multimorbidity was operationalized with chronic disease counts, a list of 24 chronic conditions obtained from the Centers for Medicare & Medicaid Services Chronic Condition Warehouse [45]. Using chronic disease counts is a common method to measure multimorbidity and is significantly related to many health outcomes [46].

Figure 1. Revised Technology Acceptance Model guiding the study.



We operationalized the independent variables (perceived usefulness and perceived ease of use) and dependent variable (intention to adopt) using adapted versions of the 3 original TAM scales [22,23,30,47], which all had good reliability (Cronbach α>.9) and validity [26,30,47]. Researchers have modified the original scales to measure the internet and mHealth apps and reported that the modified scales had good internal consistency (Cronbach α>.8) [23,48]. For this study, we changed the original wording from "[information] system" to "mHealth app" and "in my job" to "manage my chronic condition," as one's job is conceptualized as self-management [22,23]. In our sample, the intention to adopt (Cronbach α=.91), perceived usefulness (Cronbach α=.96), and perceived ease of use (Cronbach α=.91) scales all had high internal consistency.

Social influence was measured using the Social Influence Scale developed when the TAM was expanded [30]. The original scale had good reliability and validity [30], and an adapted version measuring social influence in the context of the intention to adopt mHealth apps among patients with heart failure had good internal consistency (Cronbach α=.91) [23]. In our sample, the scale had good internal consistency (Cronbach α=.78).

We measured caregiver burden using the 12-item short-form version of the Zarit Burden Interview (ZBI), which has been widely used in dementia caregiving research and found to have good internal consistency, test-retest reliability, and strong correlations with the full ZBI [49-51]. In this study sample, the ZBI instrument had Cronbach α=.90. The number of hours providing care weekly was operationalized using items from the National Long-Term Care Survey that Gitlin et al [52] shortened and adapted for use with dementia caregivers. These items ask how much time caregivers spend helping a person with dementia to perform certain activities of daily living or instrumental activities of daily living. The instrument in our sample had good internal consistency (Cronbach α=.84).

Finally, caregivers' burden of chronic disease and treatment was operationalized using the Illness Intrusiveness Ratings Scale (IIRS). This 13-item instrument measures the degree to which a disease and its treatment disrupt one's life and activities [53]. Numerous studies have validated the IIRS in various populations with a chronic disease and have supported its reliability and validity [53]. In our caregiving sample, the IIRS had excellent internal consistency (Cronbach α=.93).

Handling Fraudulent and Missing Data

Some web-based surveys were anonymous. Thus, fake or fraudulent survey responses were potential issues that could affect research integrity [54]. REDCap does not collect IP addresses or cookies. Thus, we included other methods for detecting and handling fraudulent responses. For example, we reviewed the web-based survey completion times, response patterns, participants' contact information, and contact attempts. Furthermore, the participants needed to fill out a petty cash voucher to be reimbursed for the study, which allowed the team

to verify information for some respondents; however, not all participants included in the analyses completed a voucher. Guided by the recommendations in the study by Teitcher et al [54], we excluded survey responses (14/186, 7.5%) that had (1) very short survey completion times (limits established by mock survey and average completion times), (2) unvalidated email addresses (eg, no responses to emails), and (3) inconsistent response patterns (eg, *Christmas tree* answers).

Next, we examined the data for missing, *don't know*, and *refused to answer* values. All variables had less than 4% *don't know* and 1% missing values, except for the question asking participants if they had other chronic conditions (5/117, 4.3% missing). We treated *don't know* and *refused to answer* choices as missing values and imputed a neutral or very conservative (eg, no chronic condition) value for each missing answer.

Data Analyses

Aim 1: Testing Hypothesis 1

We used descriptive statistics (mean, median, and SD) to summarize the variables and examined the distributions of independent and dependent continuous variables. We also examined the correlation matrix of bivariate associations between the independent variables and the dependent variable. All TAM variables had left-skewed distributions, with 70.1% (82/117) of the participants choosing values above neutral (somewhat agree and higher; Table S1 in Multimedia Appendix 2). We originally planned to model the outcome using linear regression; however, the data violated the assumptions of linear regression (the predicted values were associated with residual values) even after linear transformations of the outcome. Thus, we dichotomized the outcome and modeled it using multiple logistic regression.

We applied a data-driven and theoretical approach to dichotomize the TAM variables into high and low groups. Specifically, we used an approximate median split (55/45) and theoretical cutoff points for people who moderately agreed to strongly agreed that they intended to adopt mHealth apps and perceived mHealth apps as useful and easy to use. We used a similar approach for the social influence variable (people who more than somewhat agreed). The self-management and caregiver burden variables were normally distributed; thus, we dichotomized these variables at their medians. Finally, caregiving time was dichotomized into high (≥21 hours/week) and low (<21 hours/week), following a published cutoff score [55].

For hypothesis testing, each independent variable was individually regressed onto the outcome, controlling for age, gender, income, and multimorbidity, which have been associated with technology adoption in prior studies [30,56,57]. Next, any independent variables in the initial adjusted regression models with $P<.15$ were included in the final regression model [58]. We also assessed for multicollinearity in the final model, but statistics supported that multicollinearity was not an issue (average variance inflation factor=1.45).

Aim 2: Exploring Moderation

For moderation testing, we used the final model from the aim 1 analyses. Subsequently, we dichotomized race or ethnicity into White, non-Hispanic and people of color and education at its median (16 years). We created interaction terms for each dichotomized independent variable: race or ethnicity and education. All statistically significant interaction terms ($P<.05$) were included in the final model for aim 2.

Results

Sample Characteristics

The study team recruited 498 people interested in the study (Figure 2). The final sample consisted of 117 eligible caregivers; 59.8% (70/117) completed the web-based survey, and 40.1% (47/117) completed the phone survey (Table 1). Only 1 Spanish-speaking caregiver completed the Spanish web-based survey, although 79 were recruited and 11 were eligible.

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flow chart of study recruitment.

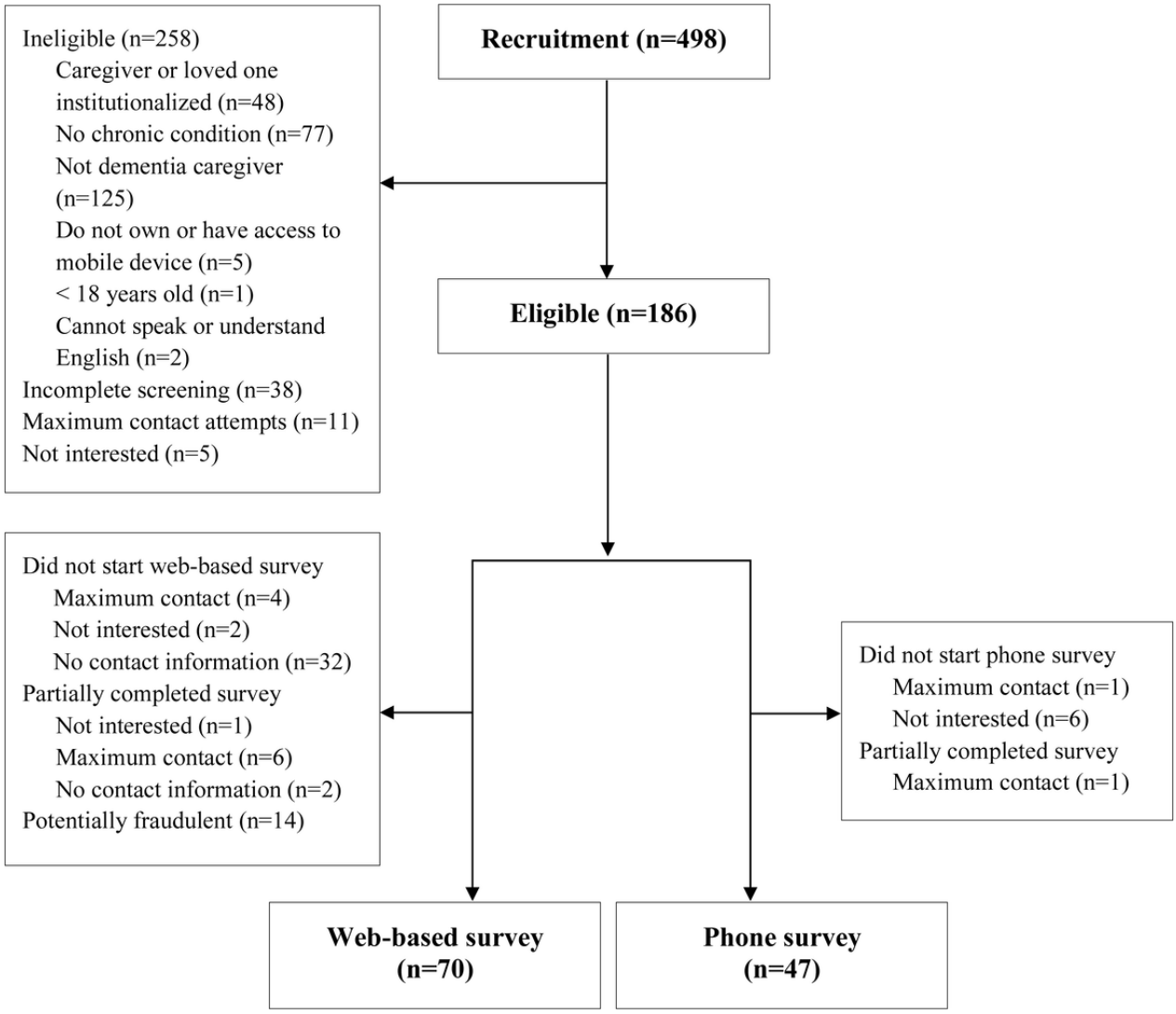


Table 1. Sociodemographic characteristics of the study sample of dementia caregivers living with a chronic health condition (N=117).

Sociodemographic characteristic	Values
Age (years), mean (SD)	52.7 (17.4)
Education (years), mean (SD)	16 (3.3)
Gender, n (%)	
Female	92 (78.6)
Male	25 (21.4)
Race or ethnicity, n (%)	
White or non-Hispanic	54 (46.2)
Black or African American	31 (26.5)
Hispanic or Latino	17 (14.5)
Asian	11 (9.4)
Native American	2 (1.7)
Multiple	2 (1.7)
Marital status, n (%)	
Married or living as married	58 (49.6)
Never married	34 (29.1)
Widowed, divorced, or separated	23 (219.7)
Refused to answer	2 (1.7)
Income (financial strain), n (%)	
Not at all or not very difficult	53 (45.3)
Somewhat difficult	48 (41)
Very or extremely difficult	16 (13.7)
Chronic health conditions[a]	
Value, mean (SD)	4 (2.5)
Common chronic conditions, n (%)	
Hypertension	55 (47)
Depression	50 (42.7)
Hyperlipidemia	40 (34.2)
Rheumatoid arthritis or osteoarthritis	39 (33.3)
Asthma	32 (27.4)
Migraine or chronic headache	32 (27.4)
Mental health condition	27 (23.1)
Diabetes (type 1 and 2)	24 (20.5)
Cataracts	23 (19.7)
Relationship to person with dementia, n (%)	
Child or child-in-law	53 (45.3)
Grandchild	22 (18.8)
Spouse or significant other	22 (18.8)
Other (family member)	12 (10.3)
Friend	8 (6.8)
Paid to provide care, n (%)	
No	106 (90.6)
Yes	10 (8.5)

Sociodemographic characteristic	Values
Refused to answer	1 (0.9)
Time spent caregiving per week (hours), mean (SD)	27.3 (30.8)

[a]Nine most common chronic conditions in the sample.

On average, the caregivers were aged approximately 53 years (SD 17.4), with an age range of 19-88 years. Most of the sample consisted of women (92/117, 78.6%), and more than half were minorities (63/117, 53.8%). Approximately half of the caregivers were married or living as married (58/117, 49.6%), and 29.1% (34/117) had never married. On average, the caregivers had completed 16 (SD 3.3) years of education, and more than half of the caregivers (64/117, 54.7%) reported that it was somewhat difficult to extremely difficult to manage on their income. Of the 117 participants, 53 (45.3%) were the child or child-in-law of the person with dementia, with an even proportion of caregivers being the spouse or significant other (22/117, 18.8%) or grandchild (22/117, 18.8%). The caregivers had, on average, 4 (SD 2.5) chronic health conditions, with a range of 1-15 (Table 1).

The web-based survey respondents were, on average, approximately 20 years younger (t_{115}=–7.81; *P*<.001), had one less chronic condition (t_{115}=–3.15; *P*=.002), provided 25 fewer hours of care per week (t_{53}=–4.07; *P*<.001), and had a 15-point higher burden of chronic disease and treatment (t_{115}=4.08; *P*<.001). In addition, a greater proportion of the web-based survey respondents were the grandchild of the person with dementia (22/70, 31% compared with 0/47, 0%; χ^2_4=20.5; *P*<.001). There were no other significant differences in the sociodemographic characteristics or main variables between the web-based and phone survey respondents.

Aim 1 Results: Testing Hypothesis 1

In bivariate associations, the intention to adopt mHealth apps was significantly associated with perceived usefulness (χ^2_1=49.8; *P*<.001), perceived ease of use (χ^2_1=28.7; *P*<.001), and social influence (χ^2_1=10.2; *P*=.002). Furthermore, perceived usefulness explained 52% of the variance in the outcome (Nagelkerke R^2=0.52). However, the caregivers' intention to adopt mHealth apps was not significantly associated with burden of chronic disease and treatment (χ^2_1=3.2; *P*=.09), caregiver burden (χ^2_1=0.6; *P*=.45), or hours spent caregiving per week (χ^2_1=1.8; *P*=.18).

After controlling for age, gender, income, and multimorbidity, we found that perceived usefulness (odds ratio [OR] 31, 95% CI 10-94; *P*<.001), perceived ease of use (OR 10.2, 95% CI 4.1-25; *P*<.001), social influence (OR 3.5, 95% CI 1.6-7.7; *P*=.002), and burden of chronic disease or treatment (OR 2.3, 95% CI 0.91-5.7; *P*=.08) were individually associated with the caregivers' intention to adopt mHealth apps with a *P*<.15, the a priori screening criteria. Caregiver burden (OR 1.3, 95% CI 0.57-2.8; *P*=.57) and hours spent caregiving per week (OR 1.7, 95% CI 0.77-3.9; *P*=.18) were not associated with the intention to adopt mHealth apps in adjusted models with *P*<.15, our a priori screening criteria, and were not included in the final aim 1 model.

The final aim 1 model is presented in Table 2. After controlling for other independent variables, only perceived usefulness was statistically significantly associated with the intention to adopt mHealth apps (OR 15, 95% CI 4.3-51; *P*<.001), although the overall model was significant (Hosmer-Lemeshow test, χ^2_8=11.5; *P*=.18). Specifically, caregivers who had high perceptions that mHealth apps were useful to their self-management had 15 higher odds of intending to adopt mHealth apps compared with those with low perceptions of mHealth apps being useful for self-management.

Table 2. Final multiple logistic regression models for aims 1 and 2.

Variable[a]	Adjusted odds ratio (95% CI)	
	Aim 1 final model[b]	Aim 2 final model[c]
Step 1: control variables		
Multimorbidity	0.87 (0.7-1.1)	0.93 (0.74-1.2)
Age (years)	1.03 (0.99-1.1)	1.03 (0.99-1.1)
Gender	0.6 (0.16-2.2)	0.41 (0.1-1.6)
Income	0.66 (0.21-2.1)	0.63 (0.16-2.4)
Step 2: independent variables		
Perceived usefulness	15 (4.3-51)[d]	23 (5.6-97)[d]
Perceived ease of use	3.1 (0.95-10)	2.4 (0.67-8.7)
Social influence	1.9 (0.64-5.5)	1.8 (0.58-5.7)
Burden of chronic disease or treatment (IIRS[e])	2.5 (0.68-9.2)	0.31 (0.038-2.5)
Step 3: interaction term		
Education	—[f]	0.24 (0.034-1.6)
IIRS × education	—	31 (2.2-430)[g]

[a]Measurement of variables is as follows (variable: measurement)—multimorbidity: chronic disease counts from the Centers for Medicare & Medicaid Services Chronic Condition Warehouse [45]; age, gender, education: questions from the US Census and other national surveys; income: Likert-type question asking, "How hard is it for you to pay for the very basics like food, housing, medical care, and heating?" [44]; perceived usefulness: Perceived Usefulness Scale modified for mHealth apps [22,23]; perceived ease of use: Perceived Ease of Use Scale modified for mHealth apps [22,23]; social influence: Social Influence Scale modified for mHealth apps [23,30]; burden of chronic disease or treatment: Illness Intrusiveness Ratings Scale [53].

[b]Aim 1 final model statistics: Hosmer-Lemeshow test, P=.18; Nagelkerke R^2=0.57.

[c]Aim 2 final model statistics: Hosmer-Lemeshow test, P=.82; Nagelkerke R^2=0.62.

[d]P<.001.

[e]IIRS: Illness Intrusiveness Ratings Scale.

[f]Variables not tested in aim 1 analyses.

[g]P=.01.

Aim 2 Results: Exploring Moderation

After exploring moderation, race or ethnicity did not significantly change the relationship between the independent variables and the outcome. The only statistically significant interaction term associated with the caregivers' intention to adopt mHealth apps was education and burden of chronic disease or treatment (OR 31, 95% CI 2.2-430; P=.01; see Tables S2 and S3 in Multimedia Appendix 2 for group comparisons).

In the final model with the interaction terms included, perceived usefulness (OR 23, 95% CI 5.6-97; P<.001) and the interaction term for education and burden of chronic disease or treatment (OR 31, 95% CI 2.2-430; P=.01) were statistically significantly associated with the intention to adopt mHealth apps. Specifically, the odds of intending to adopt an mHealth app were 23 times greater among caregivers with high beliefs that mHealth apps are useful for self-management compared with those with low beliefs that mHealth apps are useful, controlling for all other variables. In addition, the odds of intending to adopt mHealth apps for self-management were 31.6 times greater among caregivers with a high level of education and high burden of chronic disease and treatment compared with those with a low level of education and low burden of chronic disease and treatment. The other independent and control variables were not significant, although the overall model was significant (Hosmer-Lemeshow test, χ^2_8=4.4; P=.82) and explained 62% of the variance in the outcome.

Discussion

Principal Findings

The aim of this study was to explore the barriers to and facilitators of the intention to adopt mHealth apps for self-management among dementia caregivers with a chronic condition. In our study of 117 caregivers, we found that perceived usefulness explained 52% of the variance and was the strongest predictor of caregivers' intention to adopt mHealth apps for their self-management. Furthermore, after controlling for perceived usefulness, other independent variables were no longer significantly associated with the intention to adopt mHealth apps. None of the caregiving variables were significantly associated with the caregivers' intention to adopt mHealth apps in any model. We also found that caregivers with a high education level and greater burden of chronic disease and treatment had a significantly greater intention to adopt mHealth apps for their self-management than those with a low education level and low burden of chronic disease and treatment.

Perceived usefulness has consistently been a strong predictor of the intention to adopt mHealth solutions among older adults and persons with a chronic condition [24,31,59]. In later iterations of the TAM, perceived usefulness constructs had strong power to predict and were a key determinant of the intention to adopt various health technologies, including mHealth solutions [24,31,59]. For example, Dou et al [24] reported that perceived usefulness had a strong, significant positive association with the intention to adopt mHealth apps, whereas perceived ease of use was only significant when the mHealth apps were also perceived as useful. When designing mHealth app interventions for older adults with a chronic condition, it is critical to apply a user-centered design approach, which includes understanding which app features are most relevant for certain populations [60]. Future research should consider the specific features of mHealth apps that caregivers perceive as useful to their self-management, which can further facilitate their mHealth app adoption.

Although only perceived usefulness was statistically significant, perceived ease of use was clinically meaningful because caregivers who believed that mHealth apps were easy to use had 2.4 times greater intention to adopt them. This finding may not have reached statistical significance because of insufficient sample size or the age of our caregiving sample, reflecting a younger, more tech-savvy generation. For example, we only included caregivers who owned or had access to mobile devices. Previous studies have found that younger adults have higher mobile device ownership and mobile app use, as well as better technology skills than older generations [57,61]. In a larger sample of 381 dementia caregivers who were older adults (mean age 63 years, SD 13 years), Xiong et al [62] reported that ease of installation and use of caregiving-supportive technologies were the most important factors in the caregivers' decision to adopt these technologies, although the study did not investigate perceived usefulness. Other research has supported the finding that the ease of using mHealth apps is an important consideration for older adults because of cognitive, motivational, and physical barriers [60,63,64]. Nevertheless, the study by Burstein et al [29] found that after controlling for perceived usefulness, ease of use was not significantly associated with willingness to adopt caregiving-supportive technologies among older adults (mean age 59 years), similar to our findings. Taken as a whole, existing TAM research suggests that perceived usefulness is a significant facilitator of caregivers' intention to adopt technology, although ease of use may be more salient for older adult caregivers or for sustained engagement with mHealth apps, rather than for adoption [27].

In our caregiving sample, social influence had a larger, although statistically nonsignificant, OR of 1.8 (95% CI 0.58-5.7). Existing studies on the significance of social influence with regard to health-related technology adoption have been mixed. Some studies report that social influence is a significant facilitator of health-related technology adoption among general consumers [65] and patients with heart failure [23], whereas others report that it is not significant among older adults [59]. Among dementia caregivers, Dai et al [28] found that social influence significantly positively predicted the caregivers' intention to adopt wearable devices to manage the care recipient's health. However, social influence was not significantly associated with mHealth app adoption for caregivers' self-management in our sample.

This discrepancy in the findings may be related to differences in population, type of technology, or sample demographics. For example, our sample consisted of caregivers who were predominantly English-speaking, middle-aged, and the child or grandchild of the person with dementia. Compared with our caregiving sample, the sample in the study by Dai et al [28] consisted of younger caregivers, with more men, who lived in sub-Saharan Africa. In addition, our outcome investigated mHealth apps for caregivers' self-management, which is different from caregiving technologies [28]. Caregivers have a high interest in adopting caregiving technologies, but much less is known about their interest in adopting technologies for their self-management [32,66]. In-depth qualitative investigations can improve our understanding of mechanisms by which social influence impacts technology adoption; and whether social influence is more important for certain groups (eg, older caregivers) or technologies (caregiving vs self-management technologies).

In our study, social influence reflected subjective norm (perceptions that people who are important in your life believe that you should perform an action) from the Theory of Reasoned Action [23,30]. However, social support is another construct relevant to older adults' technology adoption that reflects the quality of social relationships [67,68] and may affect caregivers' adoption of mHealth apps. For example, previous qualitative studies have suggested that some dementia caregivers have poor technology literacy and rely on family to assist with using technology [68,69], although quantitative studies have found that dementia caregivers have good eHealth literacy [21]. A recent cross-sectional study supported that both subjective norm and social relationships were significant correlates of the intention to adopt mHealth apps among older adults [70]. Thus, future research is warranted to understand how social support and social influence may interact to affect mHealth app adoption among caregivers and how social support or influence may differ according to caregivers' technology literacy.

Furthermore, our sample size (n=117) was smaller than the samples in the studies by Dai et al (n=350) [28], Cajita et al (n=129) [23], and Kim and Park (n=728) [65]. Similar to the ease-of-use variable, it is possible that social influence has a smaller effect size, requiring larger samples to detect a significant relationship. Researchers should consider conducting meta-analyses to determine the effect sizes required to detect statistically significant relationships among the TAM variables. A meta-analysis will provide precise effect size estimates, with greater generalizability.

Caregiver burden and the hours spent caregiving did not contribute significantly to explaining the intention to adopt mHealth apps among family caregivers. Although these 2 caregiving factors negatively impact caregivers' self-care [5], our findings suggest that they may not be relevant to caregivers' decisions about whether to adopt mHealth apps for self-management. The median time spent caring in our sample (18.3 hours) was lower than the US population average for

dementia caregivers (26.3 hours) [1], although our sample had high levels of burden (mean ZBI score of 21), which reflects the findings of other researchers [49,71,72]. Nevertheless, our sample consisted of middle-aged and well-educated caregivers. Thus, additional research is needed to test whether these findings can be extrapolated to caregivers who are older adults and less educated.

We found that the burden of chronic disease and treatment was not significantly associated with caregivers' intention to adopt mHealth apps. Our study finding conflicts with that of existing studies. Other researchers have found that perceived disease threats were significantly associated with the intention to adopt mHealth solutions among persons with a chronic condition [24,31]. A possible explanation is that our concept and the methods we used to measure it were different. We examined the current burden of chronic disease and treatment on caregivers' lives, not their perceptions of the future consequences of a disease, as in previous studies [24,31]. Thus, it is possible that the current burden of chronic disease and treatment may not motivate the adoption of mHealth apps compared with the future perceived threats of a chronic disease. Further research is required to explore this proposition.

In our sample, the caregivers' education and burden of chronic disease and treatment interacted to produce a greater and significant effect on their intention to adopt mHealth apps. The OR (31, 95% CI 2.2-430) should be interpreted with caution because of the smaller number of caregivers in the high and low groups (Table S3 in Multimedia Appendix 2). To the best of our knowledge, very few studies have investigated how sociodemographic variables interact with chronic disease or self-management variables to affect technology adoption. A prior study investigated how age and perceived disease threat interacted to influence the intention to adopt an mHealth app and found that it was not statistically significant [31]. As prior studies have not yet examined how education and chronic disease factors may interact to affect the intention to adopt mHealth apps, additional research is needed to support this finding.

Interpreted in the context of existing research, our study offers new insights into the factors related to caregivers' intention to adopt mHealth apps for self-management. However, additional research is still needed to maximize mHealth app adoption in this population. Furthermore, the diversity of populations, mHealth strategies, and study findings substantiate the importance of user-centered design and the development of mHealth solutions with the end users as key stakeholders [60,66]. Future research should involve dementia caregivers as stakeholders throughout the process of conceptualizing, designing, and testing mHealth strategies for their self-management.

Limitations

This study has some limitations. We recruited a convenience sample using community-based (Baltimore, Maryland) and web-based methods. Thus, our results may not be generalizable to all family caregivers of people with dementia, such as those who lack access to the internet or social media. However, web-based recruitment methods enabled us to reach a larger caregiving population across the United States, which may also improve the external validity of the findings. In addition, this study was cross-sectional; thus, relationships are associative, not causal. Another limitation is that only 1 Spanish-speaking caregiver completed the survey, although 11 were eligible. We speculate that this was due to the caregivers' difficulties with navigating the REDCap survey, which does not allow researchers to change the language of prebuilt, English-only survey buttons and functionalities. The attrition of Spanish-speaking caregivers occurred when they navigated to a different part of the survey with nonmodifiable, English-only REDCap buttons. Future researchers should consider this critical limitation of the REDCap platform.

Another limitation is that the study was originally powered for linear regression. As our data violated the assumptions of linear regression, we needed to use logistic regression. This change increased the models' degrees of freedom and reduced the power to detect differences among groups. Post hoc power analyses indicated that our study had 80% power to detect an OR of 3 or higher to be statistically significant at α=.05. Thus, we may be making a type II error with some of the independent variables in our final model (such as perceived ease of use and social influence). However, in scatterplot matrices, we did not observe a linear relationship between the caregiving factors and the outcome, thus reinforcing our finding that the caregiving variables may not be relevant to caregivers' intention to adopt mHealth apps.

Conclusions

In our sample of caregivers with one or more chronic conditions, the perceived usefulness of mHealth apps was the strongest and most significant variable associated with their intention to adopt mHealth apps for self-management. Although ease of use and social influence were not statistically significant, they were clinically significant with larger ORs. Future research is needed to determine which app features are most useful for caregivers' self-management, estimate effect sizes for sample size calculations, and systematically review how relationships vary by population or type of mHealth strategy.

Our findings also support the theory that the caregiving factors may not influence caregivers' intention to adopt mHealth apps for self-management. Thus, mHealth solutions may overcome the barriers to caregivers' self-management. Furthermore, caregivers with a high education level and greater burden of chronic disease and treatment have a higher likelihood of intending to adopt mHealth apps for self-management. Future research should explore the mechanisms by which education and self-management may interact.

Engaging dementia caregivers as stakeholders throughout the process of mHealth app conception, design, and testing can promote their adoption of mHealth apps. This process of user-centered design ensures that these apps are useful and easy to use, addresses factors relevant to caregivers, and builds support systems that encourage adoption.

Acknowledgments

The authors would like to thank Chloe Kwon, MSN, RN, for helping to create the Spanish REDCap survey, recruiting participants, and collecting and cleaning data; Hannah Parks, BSN, RN, for help with Spanish translation and collecting data; and Emerald Rivers, MSN, RN, for creating content for the study Facebook page. The authors would also like to thank Welldoc BlueStar for providing pictures of their evidence-based diabetes self-management app for inclusion in the study materials.

This study was supported by grants from the National Institute of Nursing Research (F31NR018373-01A1; KJWM), National Center for Advancing Translational Sciences (TL1TR003100-01; Mary Catherine Beach), Sigma Foundation for Nursing (Sigma/Doris Bloch Research Award), and Scholl Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Authors' Contributions

KJWM designed and conducted the study, cleaned and analyzed the data, interpreted the results, and drafted the first version of the manuscript. CB led the power analyses and contributed to the data analysis and interpretation of the results. HRH assisted with designing and conducting the study, analyzing data, and interpreting the results. All authors contributed to the conception and design of the study, revised the manuscript for important intellectual content, and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CHERRIES (Checklist for Reporting Results of Internet E-Surveys) checklist.
[PDF File (Adobe PDF File), 93 KB - mhealth_v9i8e27926_app1.pdf]

Multimedia Appendix 2
Additional data on study variables and cross-tabulations for interaction term.
[PDF File (Adobe PDF File), 27 KB - mhealth_v9i8e27926_app2.pdf]

References

1. Alzheimer's Association. 2021 Alzheimer's disease facts and figures. Alzheimers Dement 2021 Mar;17(3):327-406. [doi: 10.1002/alz.12328] [Medline: 33756057]
2. Spillman B, Wolff J, Freedman V, Kasper J. Informal caregiving for older Americans: an analysis of the 2011 National Study of Caregiving. Office of the Assistant Secretary for Planning and Evaluation. 2014 Apr. URL: http://aspe.hhs.gov/daltcp/reports/2014/NHATS-IC.cfm [accessed 2020-11-03]
3. Caregiving for family and friends: a public health issue. Centers for Disease Control and Prevention. 2019 Jul 30. URL: https://www.cdc.gov/aging/caregiving/caregiver-brief.html [accessed 2020-11-03]
4. Wang XR, Robinson KM, Carter-Harris L. Prevalence of chronic illnesses and characteristics of chronically ill informal caregivers of persons with dementia. Age Ageing 2014 Jan;43(1):137-141 [FREE Full text] [doi: 10.1093/ageing/aft142] [Medline: 24072413]
5. Waligora KJ, Bahouth MN, Han H. The self-care needs and behaviors of dementia informal caregivers: a systematic review. Gerontologist 2019 Sep 17;59(5):565-583. [doi: 10.1093/geront/gny076] [Medline: 29931147]
6. Richard AA, Shea K. Delineation of self-care and associated concepts. J Nurs Scholarsh 2011 Sep;43(3):255-264. [doi: 10.1111/j.1547-5069.2011.01404.x] [Medline: 21884371]
7. Acton GJ. Health-promoting self-care in family caregivers. West J Nurs Res 2002 Feb;24(1):73-86. [doi: 10.1177/01939450222045716] [Medline: 11829276]
8. Capistrant BD, Moon JR, Berkman LF, Glymour MM. Current and long-term spousal caregiving and onset of cardiovascular disease. J Epidemiol Community Health 2012 Oct;66(10):951-956 [FREE Full text] [doi: 10.1136/jech-2011-200040] [Medline: 22080816]
9. Oliveira D, Zarit SH, Orrell M. Health-promoting self-care in family caregivers of people with dementia: the views of multiple stakeholders. Gerontologist 2019 Sep 17;59(5):501-511. [doi: 10.1093/geront/gnz029] [Medline: 30953585]
10. Pinquart M, Sörensen S. Differences between caregivers and noncaregivers in psychological health and physical health: a meta-analysis. Psychol Aging 2003 Jun;18(2):250-267. [doi: 10.1037/0882-7974.18.2.250] [Medline: 12825775]
11. Gallant MP, Connell CM. Predictors of decreased self-care among spouse caregivers of older adults with dementing illnesses. J Aging Health 1997 Aug;9(3):373-395. [doi: 10.1177/089826439700900306] [Medline: 10182399]
12. Wang X, Robinson KM, Hardin HK. The impact of caregiving on caregivers' medication adherence and appointment keeping. West J Nurs Res 2015 Dec;37(12):1548-1562. [doi: 10.1177/0193945914533158] [Medline: 24807892]

13. Giebel C, Hanna K, Cannon J, Eley R, Tetlow H, Gaughan A, et al. Decision-making for receiving paid home care for dementia in the time of COVID-19: a qualitative study. BMC Geriatr 2020 Sep 09;20(1):333 [FREE Full text] [doi: 10.1186/s12877-020-01719-0] [Medline: 32900360]
14. Ng KY, Zhou S, Tan SH, Ishak ND, Goh ZZ, Chua ZY, et al. Understanding the psychological impact of COVID-19 pandemic on patients with cancer, their caregivers, and health care workers in Singapore. JCO Glob Oncol 2020 Oct;6:1494-1509 [FREE Full text] [doi: 10.1200/GO.20.00374] [Medline: 33017179]
15. Park SS. Caregivers' mental health and somatic symptoms during COVID-19. J Gerontol B Psychol Sci Soc Sci 2021 Mar 14;76(4):235-240 [FREE Full text] [doi: 10.1093/geronb/gbaa121] [Medline: 32738144]
16. Lecomte T, Potvin S, Corbière M, Guay S, Samson C, Cloutier B, et al. Mobile apps for mental health issues: meta-review of meta-analyses. JMIR Mhealth Uhealth 2020 May 29;8(5):e17458 [FREE Full text] [doi: 10.2196/17458] [Medline: 32348289]
17. Osborn J, Ajakaiye A, Cooksley T, Subbe CP. Do mHealth applications improve clinical outcomes of patients with cancer? A critical appraisal of the peer-reviewed literature. Support Care Cancer 2020 Mar;28(3):1469-1479 [FREE Full text] [doi: 10.1007/s00520-019-04945-4] [Medline: 31273501]
18. Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med 2020;3:4 [FREE Full text] [doi: 10.1038/s41746-019-0206-x] [Medline: 31970289]
19. Wang Y, Min J, Khuri J, Xue H, Xie B, Kaminsky LA, et al. Effectiveness of mobile health interventions on diabetes and obesity treatment and management: systematic review of systematic reviews. JMIR Mhealth Uhealth 2020 Apr 28;8(4):e15400 [FREE Full text] [doi: 10.2196/15400] [Medline: 32343253]
20. Kim H, Powell MP, Bhuyan SS. Seeking medical information using mobile apps and the internet: are family caregivers different from the general public? J Med Syst 2017 Mar;41(3):38. [doi: 10.1007/s10916-017-0684-9] [Medline: 28101781]
21. Rathnayake S, Moyle W, Jones C, Calleja P. Family carers' needs related to management of functional disability in dementia care and use of mHealth applications in health information seeking: an online survey. Collegian 2020 Jun;27(3):288-297. [doi: 10.1016/j.colegn.2019.09.001]
22. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
23. Cajita MI, Hodgson NA, Budhathoki C, Han H. Intention to use mhealth in older adults with heart failure. J Cardiovasc Nurs 2017 Feb 28;32(6):1-7 [FREE Full text] [doi: 10.1097/JCN.0000000000000401] [Medline: 28248747]
24. Dou K, Yu P, Deng N, Liu F, Guan Y, Li Z, et al. Patients' acceptance of smartphone health technology for chronic disease management: a theoretical model and empirical test. JMIR Mhealth Uhealth 2017 Dec 06;5(12):e177 [FREE Full text] [doi: 10.2196/mhealth.7886] [Medline: 29212629]
25. Davis F. A technology acceptance model for empirically testing new end-user information systems: theory and results. Thesis (Ph. D.) - Massachusetts Institute of Technology, Sloan School of Management. 1986. URL: http://dspace.mit.edu/handle/1721.1/15192 [accessed 2021-07-13]
26. Davis F, Bagozzi R, Warshaw P. User acceptance of computer technology: a comparison of two theoretical models. Manag Sci 1989 Aug;35(8):982-1003 [FREE Full text] [doi: 10.1287/mnsc.35.8.982]
27. Birkhoff SD, Smeltzer SC. Perceptions of smartphone user-centered mobile health tracking apps across various chronic illness populations: an integrative review. J Nurs Scholarsh 2017 Jul;49(4):371-378. [doi: 10.1111/jnu.12298] [Medline: 28605151]
28. Dai B, Larnyo E, Tetteh EA, Aboagye AK, Musah AI. Factors affecting caregivers' acceptance of the use of wearable devices by patients with dementia: an extension of the unified theory of acceptance and use of technology model. Am J Alzheimers Dis Other Demen 2020;35:1533317519883493 [FREE Full text] [doi: 10.1177/1533317519883493] [Medline: 31679390]
29. Burstein A, DaDalt O, Kramer B, D'Ambrosio L, Coughlin J. Dementia caregivers and technology acceptance: interest outstrips awareness. Gerontechnology 2015;14(1):45-56. [doi: 10.4017/gt.2015.14.1.005.00]
30. 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]
31. Zhu Z, Liu Y, Che X, Chen X. Moderating factors influencing adoption of a mobile chronic disease management system in China. Inform Health Soc Care 2018 Jan;43(1):22-41. [doi: 10.1080/17538157.2016.1255631] [Medline: 28068149]
32. Ghahramani F, Wang J. Intention to adopt mhealth apps among informal caregivers: cross-sectional study. JMIR Mhealth Uhealth 2021 Mar 17;9(3):e24755 [FREE Full text] [doi: 10.2196/24755] [Medline: 33729166]
33. Frisbee KL. Variations in the use of mHealth tools: the VA mobile health study. JMIR Mhealth Uhealth 2016 Jul 19;4(3):e89 [FREE Full text] [doi: 10.2196/mhealth.3726] [Medline: 27436165]
34. Piette JD, Striplin D, Marinec N, Chen J, Aikens JE. A randomized trial of mobile health support for heart failure patients and their informal caregivers: impacts on caregiver-reported outcomes. Med Care 2015 Aug;53(8):692-699 [FREE Full text] [doi: 10.1097/MLR.0000000000000378] [Medline: 26125415]
35. Mobile fact sheet. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2021-01-14]

36. Ryan C. Computer and internet use in the United States: 2016. American Community Survey Reports - ACS-39. 2018 Aug. URL: https://www.census.gov/content/dam/Census/library/publications/2018/acs/ACS-39.pdf [accessed 2021-01-14]
37. Hamilton EC, Saiyed F, Miller CC, Eguia A, Fonseca AC, Baum GP, et al. The digital divide in adoption and use of mobile health technology among caregivers of pediatric surgery patients. J Pediatr Surg 2018 Aug;53(8):1478-1493. [doi: 10.1016/j.jpedsurg.2017.08.023] [Medline: 28927983]
38. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
39. Patridge EF, Bardyn TP. Research Electronic Data Capture (REDCap). J Med Libr Assoc 2018 Jan 12;106(1):142-144. [doi: 10.5195/jmla.2018.319]
40. Desveaux L, Shaw J, Saragosa M, Soobiah C, Marani H, Hensel J, et al. A mobile app to improve self-management of individuals with type 2 diabetes: qualitative realist evaluation. J Med Internet Res 2018 Mar 16;20(3):e81 [FREE Full text] [doi: 10.2196/jmir.8712] [Medline: 29549070]
41. Quinn CC, Shardell MD, Terrin ML, Barr EA, Ballew SH, Gruber-Baldini AL. Cluster-randomized trial of a mobile phone personalized behavioral intervention for blood glucose control. Diabetes Care 2011 Sep;34(9):1934-1942 [FREE Full text] [doi: 10.2337/dc11-0366] [Medline: 21788632]
42. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
43. Mendez KJ, Han H. Recruiting racially and ethnically diverse dementia family caregivers for survey research. Innov Aging 2020;4(Suppl 1):353. [doi: 10.1093/geroni/igaa057.1135]
44. Puterman E, Haritatos J, Adler NE, Sidney S, Schwartz JE, Epel ES. Indirect effect of financial strain on daily cortisol output through daily negative to positive affect index in the Coronary Artery Risk Development in Young Adults Study. Psychoneuroendocrinology 2013 Dec;38(12):2883-2889 [FREE Full text] [doi: 10.1016/j.psyneuen.2013.07.016] [Medline: 23969421]
45. Condition categories. Chronic Conditions Data Warehouse. URL: https://www.ccwdata.org/web/guest/condition-categories [accessed 2017-09-13]
46. Huntley AL, Johnson R, Purdy S, Valderas JM, Salisbury C. Measures of multimorbidity and morbidity burden for use in primary care and community settings: a systematic review and guide. Ann Fam Med 2012;10(2):134-141 [FREE Full text] [doi: 10.1370/afm.1363] [Medline: 22412005]
47. Venkatesh V, Davis F. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manag Sci 2000 Feb;46(2):186-204 [FREE Full text] [doi: 10.1287/mnsc.46.2.186.11926]
48. Tung F, Chang S, Chou C. An extension of trust and TAM model with IDT in the adoption of the electronic logistics information system in HIS in the medical industry. Int J Med Inform 2008 May;77(5):324-335. [doi: 10.1016/j.ijmedinf.2007.06.006] [Medline: 17644029]
49. Bédard M, Molloy DW, Squire L, Dubois S, Lever JA, O'Donnell M. The Zarit Burden Interview: a new short version and screening version. Gerontologist 2001 Oct;41(5):652-657. [doi: 10.1093/geront/41.5.652] [Medline: 11574710]
50. Seng BK, Luo N, Ng WY, Lim J, Chionh HL, Goh J, et al. Validity and reliability of the Zarit Burden Interview in assessing caregiving burden. Ann Acad Med Singap 2010 Oct;39(10):758-763. [Medline: 21063635]
51. Vitaliano PP, Young HM, Russo J. Burden: a review of measures used among caregivers of individuals with dementia. Gerontologist 1991 Feb;31(1):67-75. [doi: 10.1093/geront/31.1.67] [Medline: 2007476]
52. Gitlin LN, Roth DL, Burgio LD, Loewenstein DA, Winter L, Nichols L, et al. Caregiver appraisals of functional dependence in individuals with dementia and associated caregiver upset: psychometric properties of a new scale and response patterns by caregiver and care recipient characteristics. J Aging Health 2005 Apr;17(2):148-171 [FREE Full text] [doi: 10.1177/0898264304274184] [Medline: 15750049]
53. Devins GM. Using the illness intrusiveness ratings scale to understand health-related quality of life in chronic disease. J Psychosom Res 2010 Jun;68(6):591-602. [doi: 10.1016/j.jpsychores.2009.05.006] [Medline: 20488277]
54. Teitcher JE, Bockting WO, Bauermeister JA, Hoefer CJ, Miner MH, Klitzman RL. Detecting, preventing, and responding to "fraudsters" in internet research: ethics and tradeoffs. J Law Med Ethics 2015;43(1):116-133 [FREE Full text] [doi: 10.1111/jlme.12200] [Medline: 25846043]
55. Process evaluation of the Older Americans Act Title III-E National Family Caregiver Support Program: final report. The Lewin Group, Inc. 2016 Mar. URL: https://acl.gov/sites/default/files/programs/2017-02/NFCSP_Final_Report-update.pdf [accessed 2020-11-02]
56. Kim H. Understanding internet use among dementia caregivers: results of secondary data analysis using the US caregiver survey data. Interact J Med Res 2015 Feb 23;4(1):27-38 [FREE Full text] [doi: 10.2196/ijmr.3127] [Medline: 25707033]
57. Jiang J, Zhu Q, Zheng Y, Zhu Y, Li Y, Huo Y. Perceptions and acceptance of mhealth in patients with cardiovascular diseases: a cross-sectional study. JMIR Mhealth Uhealth 2019 Feb 04;7(2):e10117 [FREE Full text] [doi: 10.2196/10117] [Medline: 30714942]
58. Tabachnick B, Fidell L. Using Multivariate Statistics, 5th Ed. Boston, MA: Pearson Education; 2007:1-983.

59. Kavandi H, Jaana M. Factors that affect health information technology adoption by seniors: a systematic review. Health Soc Care Community 2020 Nov;28(6):1827-1842. [doi: 10.1111/hsc.13011] [Medline: 32378769]
60. Matthew-Maich N, Harris L, Ploeg J, Markle-Reid M, Valaitis R, Ibrahim S, et al. Designing, implementing, and evaluating mobile health technologies for managing chronic conditions in older adults: a scoping review. JMIR Mhealth Uhealth 2016 Jun 09;4(2):e29 [FREE Full text] [doi: 10.2196/mhealth.5127] [Medline: 27282195]
61. Gallagher R, Roach K, Sadler L, Glinatsis H, Belshaw J, Kirkness A, et al. Mobile technology use across age groups in patients eligible for cardiac rehabilitation: survey study. JMIR Mhealth Uhealth 2017 Oct 24;5(10):e161 [FREE Full text] [doi: 10.2196/mhealth.8352] [Medline: 29066425]
62. Xiong C, Ye B, Mihailidis A, Cameron JI, Astell A, Nalder E, et al. Sex and gender differences in technology needs and preferences among informal caregivers of persons with dementia. BMC Geriatr 2020 May 18;20(1):176 [FREE Full text] [doi: 10.1186/s12877-020-01548-1] [Medline: 32423406]
63. Cajita MI, Hodgson NA, Lam KW, Yoo S, Han H. Facilitators of and barriers to mHealth adoption in older adults with heart failure. Comput Inform Nurs 2018 Aug;36(8):376-382. [doi: 10.1097/CIN.0000000000000442] [Medline: 29742549]
64. Wildenbos GA, Jaspers MW, Schijven MP, Dusseljee-Peute LW. Mobile health for older adult patients: using an aging barriers framework to classify usability problems. Int J Med Inform 2019 Dec;124:68-77 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.01.006] [Medline: 30784429]
65. Kim J, Park H. Development of a health information technology acceptance model using consumers' health behavior intention. J Med Internet Res 2012;14(5):e133 [FREE Full text] [doi: 10.2196/jmir.2143] [Medline: 23026508]
66. AARP. Caregivers and technology: what they want and need. Am Asso Retir Person 2016 Apr:1-48 [FREE Full text] [doi: 10.26419/res.00191.002]
67. Reynolds A, Osborne T. Chapter 2: Promoting technology adoption engagement in aging. In: Chau D, Osborne TF, editors. Using Technology to Improve Care of Older Adults. New York, NY: Springer Publishing Company; 2018:1-250.
68. Ruggiano N, Brown EL, Shaw S, Geldmacher D, Clarke P, Hristidis V, et al. The potential of information technology to navigate caregiving systems: perspectives from dementia caregivers. J Gerontol Soc Work 2019;62(4):432-450. [doi: 10.1080/01634372.2018.1546786] [Medline: 30422754]
69. Rathnayake S, Jones C, Calleja P, Moyle W. Family carers' perspectives of managing activities of daily living and use of mHealth applications in dementia care: a qualitative study. J Clin Nurs 2019 Dec;28(23-24):4460-4470. [doi: 10.1111/jocn.15030] [Medline: 31408554]
70. Askari M, Klaver NS, van Gestel TJ, van de Klundert J. Intention to use medical apps among older adults in The Netherlands: cross-sectional study. J Med Internet Res 2020 Sep 04;22(9):e18080 [FREE Full text] [doi: 10.2196/18080] [Medline: 32624465]
71. Rodríguez-González A, Rodríguez-Míguez E. A meta-analysis of the association between caregiver burden and the dependent's illness. J Women Aging 2020;32(2):220-235. [doi: 10.1080/08952841.2019.1700728] [Medline: 31818233]
72. Yu J, Yap P, Liew TM. The optimal short version of the Zarit Burden Interview for dementia caregivers: diagnostic utility and externally validated cutoffs. Aging Ment Health 2019 Jun;23(6):706-710. [doi: 10.1080/13607863.2018.1450841] [Medline: 29553806]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
IIRS: Illness Intrusiveness Ratings Scale
IRB: Institutional Review Board
OR: odds ratio
REDCap: Research Electronic Data Capture
TAM: Technology Acceptance Model
ZBI: Zarit Burden Interview

Edited by L Buis; submitted 13.02.21; peer-reviewed by MI Cajita, W Tang; comments to author 16.03.21; revised version received 26.04.21; accepted 04.06.21; published 31.08.21.

Please cite as:
Mendez KJW, Budhathoki C, Labrique AB, Sadak T, Tanner EK, Han HR
Factors Associated With Intention to Adopt mHealth Apps Among Dementia Caregivers With a Chronic Condition: Cross-sectional, Correlational Study
JMIR Mhealth Uhealth 2021;9(8):e27926
URL: https://mhealth.jmir.org/2021/8/e27926
doi:10.2196/27926
PMID:34463637

Original Paper

Mobile Electronic Devices as Means of Facilitating Patient Activation and Health Professional Empowerment Related to Information Seeking on Chronic Conditions and Medications: Qualitative Study

Hyllore Imeri[1,2], MPharm; Shane Desselle[3], PhD; Dardan Hetemi[1], PhD; Kreshnik Hoti[1], PhD

[1]Division of Pharmacy, Faculty of Medicine, University of Prishtina, Pristina, Kosovo

[2]Department of Pharmacy, University of Mississippi, Oxford, MS, United States

[3]Department of Pharmacy, Touro University California, Vallejo, CA, United States

Corresponding Author:
Kreshnik Hoti, PhD
Division of Pharmacy
Faculty of Medicine
University of Prishtina
George Bush Street #31
Pristina, 10000
Kosovo
Phone: 383 44945173
Email: kreshnik.hoti@uni-pr.edu

Abstract

Background: Patient activation has an impact on the management of patients' health, clinical outcomes, and treatment costs. Mobile electronic devices (MEDs) have shown the potential to engage patients in wellness behavior. Furthermore, the potentially positive role of MEDs is evident in supporting health professionals in their practice.

Objective: This study aims to explore the impact of MEDs on patient activation to search for information on chronic conditions and medications and the impact of MEDs on the empowerment of health professionals or future health professionals.

Methods: We conducted 6 focus groups—2 with health sciences students, 2 with health professionals, and 2 with hospitalized patients with chronic conditions. A protocol comprising eight questions was used to guide discussions. Audio-recorded data were transcribed verbatim and analyzed thematically; a ranking system was used to analyze the relevance of identified themes and subthemes, using a coding system depicted by the + symbol, to indicate different relevance levels.

Results: Our results suggest that MEDs can positively affect patient activation to search for chronic conditions and medication information by facilitating patients' information-seeking behavior. Key drivers leading to patients' activation to seek information related to chronic conditions and medications through MEDs were the accessibility and abundance of available and detailed information, reduced search time, information updates, and convenience in finding information at any time and place. The lack of accurate information in one's native language, access to incorrect information, and limited access to the internet were key obstacles to seeking information related to chronic conditions and medications via MEDs. In addition, findings of this study suggest that MEDs in general and mobile apps, in particular, may have a positive impact on the work routine of health care professionals as they enable them to make quicker decisions by accessing the required information faster, thus improving practice efficiency. Furthermore, the appropriate usage of MEDs by patients for seeking information about their chronic conditions and medications may positively impact the physician-patient relationship. All focus groups recognized the questionable reliability of health information on the internet and its potential negative effects on patients. Therefore, our findings suggest the need for an additional role of health professionals in assisting patients in using MEDs to search for health and medication information, such as providing reliable websites and mobile apps where patients can safely search for health-related information on the web.

Conclusions: The use of MEDs may help activate patients to seek chronic conditions and medication-related information, potentially leading to better management of their chronic conditions and medications. Our findings also highlight the positive impact MEDs may have on empowering health professionals in their practice and the need for health professionals to help patients through specific education that addresses MEDs utilization for chronic conditions and medication information seeking.

(JMIR Mhealth Uhealth 2021;9(8):e26300) doi:10.2196/26300

KEYWORDS

patient activation; mobile electronic devices; health professionals; chronic conditions; medications

Introduction

Background

Patient activation is a behavioral concept defined as "the individual's knowledge, skills and belief in managing his/her health and healthcare" [1]. Patient activation enables an understanding of why some patients engage and are actively involved in their health whereas others are not. Considering the relevance of patients' roles, health care settings have moved to a more patient-centered model, where patients are encouraged to be effective managers of their health care [2,3]. A higher patient activation level is associated with a wide range of better health outcomes [4], health-related behaviors, and health care costs [5]. In addition, although nearly half of patients assess information about their medical condition or treatments on the web [6], evidence suggests that activated persons are more likely to use web-based health information [7].

Although there have been rapid developments in mobile health (mHealth) apps, their effects on improving health and health care remain unclear [8,9]. However, the role of mobile electronic devices (MEDs) concerning patient engagement and facilitation of communication between patients and health professionals has been recognized. In this regard, several programs using mobile devices to engage and educate patients have been designed [8,10]. Evidence suggests that mobile phone apps can facilitate medication adherence by using functions such as reminder alerts and providing access to drug information [11]. Furthermore, many mobile apps to support health care professionals during their practice have been developed [12-14], thus promoting personalized and efficient health care. Therefore, the potential of MEDs, including laptops, tablets, and especially mobile phones, related to facilitating activation to navigate health information on the web is expected to be highly advantageous to not only patients but also health professionals.

Patient activation levels may increase with time, making patient activation an important focus of interventions to improve health behaviors and outcomes [6,7]. There are four patient activation levels that patients go through as they become more activated, from being disengaged and overwhelmed to maintaining positive health-related behaviors and pushing further [1]. Providing health information to patients is one of the initial steps of engagement in health care self-management. Patient education through health-related informative letters has been found to drive patient activation in patients with chronic conditions such as hypertension, suggesting a significant role of health information in patient activation [15]. However, considering the increased tendency to seek health information on the web [6] and the high rates of MED ownership [16], a potential role of MEDs in facilitating the process of patient education, and consequently the overall patient activation, could be suggested.

Objectives

Currently, there is limited evidence exploring how MEDs could facilitate the activation of patients seeking information on the web about their health and medications and the acceptance of MEDs on health professionals' work. This study aims to explore (1) the role of MEDs on patient activation toward seeking information about chronic conditions and medications and (2) the impact of MEDs on health professionals or future health professionals' empowerment.

Methods

Study Design

This qualitative study used focus groups (FGs) as a data collection method, which allowed us to explore various perspectives. Approvals to conduct the study and access to patients were provided by the ethics committee of the Faculty of Medicine, University of Pristina, and the University Clinical Center of Kosovo (UCCK) office for personal data protection, respectively.

Study Setting and FG Participants

In total, 6 FGs were organized with 4 to 7 participants. These numbers were chosen based on the literature suggesting that themes saturation can be achieved on this basis [17]. In particular, saturation was sought despite seeking input from various groups involved in the medication use process, including patients, health professionals, and student health professionals. To examine the role of MEDs in patient activation toward health information seeking from various perspectives, we conducted FGs with patients, health professionals, and future health professionals. Health sciences students were selected for FG inclusion because of their likelihood of being well-versed and high MEDs users, thus presenting the potential to better understand the acceptance of MEDs from future health professionals. Of the 6 FGs, 2 (33%) were conducted with health science students at the University of Pristina, 2 (33%) with health professionals working at the UCCK, and 2 (33%) with patients with chronic conditions who were hospitalized at the UCCK and receiving more than one medication. FGs were organized on the premises of the UCCK and the Faculty of Medicine, University of Pristina.

Recruitment

A purposive selection strategy was used, targeting health professional staff of the UCCK, patients with chronic conditions receiving health care treatment at the UCCK, and health sciences students. The study's setting in Kosovo presents an ideal environment in which to conduct the study, given the high use of internet technology and connectivity of Kosovo's population. However, there is lack of data that would suggest that Kosovo's residents leveraged this use into active involvement into technologies that potentially assists them with management of their health conditions or seeking information for health

maintenance purposes. Students were recruited using snowball sampling, whereas health professional staff and patients were approached by the facilitator of FGs and the principal author on the UCCK premises and invited to participate in the study. Before inviting patients and health professionals to participate in the study, researchers informed the respective clinics' chiefs of staff. Patients were approached during the optimal time of the day after the physicians' consultations, and FGs were conducted on the same day. Health professionals chose their optimal time and date, which was before starting their shifts or during breaks. Potential subjects who agreed to participate also received an information letter and a letter of consent as an invitation to be a participant in the FG. The design of the FG structure and protocol was based on data from previous studies [18].

FG Discussion Guide

The development of the question protocol for this study considered the concept of patient activation and the utility of MEDs in seeking health information on the web relevant to patients' health care management. In addition, the researchers who conducted the FGs clarified the definition of MEDs verbally as part of the opening statements by taking examples of mobile phones, tablets, and laptops. The question protocol consisted of an opening question, four transitory questions, and three key questions. The opening question was related to searching for health and medication information and participants' access to MED use. Transitory questions were related to difficulties in seeking information, factors that would help overcome these obstacles, and patient views on health professionals' roles. Key questions were related to the potential drivers leading to patient activation to seek information on chronic conditions and medications and the role of MEDs during this process.

The FGs were conducted by a facilitator. The principal author also attended the meetings, took notes, and audio recorded to further facilitate data analysis and mitigate against the potential bias introduced by the facilitator. All FGs were conducted in the native language of the participants, which is Albanian, and were held between May and June 2018. Audio-recorded data from FG meetings were transcribed verbatim into Microsoft Word version 2013 and translated and reviewed by 3 researchers.

Qualitative Analysis

Data were analyzed thematically in Microsoft Word, using the open, axial, and selective coding strategy [19], initially by one of the researchers of this study, and a second analysis and review were conducted by the other researcher of the study. A ranking system was used to analyze the relevance of the identified themes and subthemes. This approach has been previously reported in the literature [20]. If a similar comment was repeated for a given issue, the + symbol was used. If a comment was repeated only within 1 FG, it was marked +, if a comment was repeated in 2 to 3 FGs, it was marked as ++, and when a comment was repeated in all FGs, it was marked with +++, where + indicates low relevance, ++ indicates average relevance, and +++ indicates high relevance.

Results

Overview

A total of 6 FGs involving 31 participants were conducted on the UCCK and Faculty of Medicine premises in Pristina, Kosovo. The 2 FGs with health care professionals were conducted with the health care professional staff of the Clinic of Nephrology and the Infectious Diseases Clinic at the UCCK, 4 and 7 participants, respectively, with ages ranging from 30 to 55 years, of which 73% (8/11) were female. The 2 FGs with patients were conducted with patients admitted to either the Dermatology Clinic (UCCK) or Clinic of Hematology (UCCK), with 4 participants per FG. Patient age ranged from 45 to 75 years, of which 63% (5/8) were women. In addition, a total of 12 students agreed to participate. They were divided into 2 FGs, each composed of 6 participants, with ages ranging from 20 to 25 years, and equal participation of both men and women. The duration of each FG meeting was approximately 50 minutes.

In FG discussions, five main themes were highlighted referring to patient activation via MEDs to seek information on chronic conditions and medications. These themes pertained to motives for seeking information on chronic conditions and medications via MEDs, difficulties and obstacles to seeking information on chronic conditions and medications via MEDs, the overall activation level of patients, the impact of MEDs in activating patients to seek health-related information, and the role of health professionals in facilitating the use of MEDs to enhance patient activation toward seeking health-related information. The themes have been described in more detail below.

Motives for Seeking Information on Chronic Conditions and Medications via MEDs

For most students and health care professionals, the internet was the main source of information on chronic conditions and medication, whereas health care professionals were the main source of information for most patients. Key identified motives for using MEDs to seek health-related information were using time efficiently, reducing information ambiguity, and getting the most up-to-date medication information. The subthemes with corresponding comments are shown in Textbox 1.

Textbox 1. Theme 1—summary of comments corresponding to each of the subthemes.

Detailed Medication Information (+++)

- "I use the applications mostly for medicines, about the effect of a herb, the action mechanism and the contraindications." [student]
- "We use MED more about drugs than for diseases because it is a problem to find information and read about diseases on Google." [health professional]
- "Yes, I even use Facebook. If there is any diet to lose weight, I use it. Even when I receive medication information, I get informed, and when I go to the doctor, I ask him about what I have read, without the doctor's advice I do not use anything." [patient]

Ease and Time Efficiency of Information Access (++)

- "I think the reason why we use phones, or the Internet is often the time...you will immediately have thousands of publications or links that send you directly to the requested information. The time of searching for a particular problem is shortened, so it is a way of searching much faster than by searching in books." [student]
- "The use of MED relates to the comfort offered to find information at any time and in any place." [health professional]
- "We search drugs (online) because it is faster." [patient]

Reducing Information Ambiguity (++)

- "Yes, it has become essential to have a mobile with us. In case I encounter ambiguity..." [student]
- "Yes, we use MED, but it is a bit of a problem to open the phone directly with the patient; then the patient perceives us as we do not know things, but we tell them that, eg, a drug could have 50 or 60 commercial names..." [health professional]

New Drugs on the Market (++)

- "...when I hear a new drug name, I search it at least to have the basic information." [student]
- "We use MED for health and medicines, but mostly to get information about a certain drug that comes out in the market." [health professional]

Getting Up-to-Date Information (++)

- "...information via mobile devices, gets the information fast and uses information that is more updated than books." [student]
- "...the reasons are to be updated for a certain new medication, to recall and recapture things that could have gone through in the second plan..." [health professional]

Wider Range of Information (++)

- "The main motive of looking for a drug is the interest in knowing more drugs because the basic literature is not sufficient." [student]
- "Internet search through MED provides a wider range of information, all areas are there." [health professional]

Queries From Other Family Members or Society (+)

- "Usually a certain medical condition that I have or someone in the family does, this is the key to pushing me to research, then also for faculty issues if I need something..." [student]

Medical Condition (++)

- "On internet you can read something superficially, for illnesses we rather read in books." [health professional]
- "Our medical condition pushes us to search for information." [patient]

Difficulties and Obstacles to Seeking Information on Chronic Conditions and Medications via MEDs

Unlike participants from patient-based FGs who expressed confidence in using MEDs for information retrieval, students and health professionals reported difficulties and obstacles while seeking health and medication information. However, it should be noted that many participants in patient-based FGs did not use MEDs for information searches specifically related to health and medications. The following key difficulties were identified: lack of accurate information in the native language, limited access to the internet, and access to incorrect information. The subthemes with the illustrated comments are shown in Textbox 2.

Textbox 2. Theme 2—summary of comments corresponding to each of the subthemes.

Lack of Accurate Information in the Native Language (++)

- "If the literature was in Albanian, it might have been much easier to understand, but there are data that are not electronically in Albanian, so in other languages it is more difficult to understand." [student]
- "There is not much information in our [Albanian] language." [health professional]

Inability to Access Scientific Journals Due to Subscriptions (+)

- "The problem is the inability to access publications. In many cases, if we want to read any publication, we read only the abstract." [student]

Inability to Access the Internet (+++)

- "Another obstacle I see is access, connection to the Internet. There are quite accurate apps, but you cannot access them offline." [student]
- "Another difficulty is the Internet access to our clinics, when we spent our mobile data then our internet access ends." [health professional]
- "I do not have an internet connection." [patient]

Access to Incorrect Information (++)

- "...if we look for information in electronic books, then I do not see any difficulty, but if I search for information across different web pages, then it could be incorrect." [student]
- "There is information, but when it comes to different websites, you can read what people who are not competent for that matter have also written." [health professional]

Overcoming Obstacles in Seeking Information on Health and Medication (++)

- "Initially, it would be good to have English language knowledge as most of the accurate and up-to-date information is in English, and this would help us to research more." [student]
- "Free internet access should be provided in all clinics of UCCK." [health professional]
- "We should have consultations with pharmacists of the clinics more often because we hardly see them. In this way, we always have to find the information by ourselves, to research it online or to consult with other doctors." [health professional]

The Overall Activation Level of Patients

All participants supported the approach that, in general, it is the patients' responsibility to actively engage in their health management and, therefore, achieve a higher level of activation. This was also the case for the use of MEDs to facilitate patient activation. According to them, the physician's responsibility is in diagnosing and prescribing, whereas before and after this, it depends on the patient's behaviors. However, some patients declared that they were not the only ones responsible for actively engaging in their health management. They see themselves and the physician at the same level of responsibility. The subthemes and related comments are listed in Textbox 3.

Textbox 3. Theme 3—summary of comments that correspond to each of the subthemes.

Self-care Suggesting a High Level of Activation (+++)

- "The patient is responsible for his/her health, the doctors are also [responsible], but secondary [compared to the patient]. I think that the patient is more responsible for their own health." [student]
- "Responsibility is of the doctor when prescribing the drug, but the patient must adhere to the doctor's counsel. As far as I prescribe metformin and the patient eats baklava [sweets], then it is no longer my responsibility." [health professional]
- "For your health, you have to be responsible. If you are not responsible for yourself, then the doctor cannot be either." [patient]
- "Much depends on the patient because the doctor cannot go to the patient's home and take care of him." [patient]

Self-care Suggesting a Low Level of Activation (+)

- "We face a lot of health-related neglect by patients themselves. It is unimaginable how little we care about ourselves; we care more about our cars..." [health professional]
- "Together, if you do not help the doctor, he finds it difficult [to manage your health]." [patient]

The Impact of MEDs in Activating Patients to Seek Health-Related Information

The MEDs' impact on patients was related primarily to their empowerment by facilitating patient activation to seek health-related information. The impact was considered positive if the information was searched adequately and access to accurate and credible information was provided. However, a potential disadvantage to using MEDs from the patients' perspective, was the possibility of bypassing the physician's

visit. Health professionals suggested that when patients are more informed about their health condition, this facilitates health professionals' work because of better patient understanding and communication level. To ensure the overall positive impact of MEDs, health professionals suggested the option of providing patients with access to credible health information. Furthermore, patients can be given access to websites with health professionals available to respond to their queries. Information uncertainty and the possibility of information misinterpretation were emphasized to be dangerous. Considering this, it was suggested that patients be informed but not decide about their health based solely on the information they find on the internet. The related subthemes and corresponding comments are listed in Textbox 4.

Textbox 4. Theme 4—summary of comments corresponding to each of the subthemes.

Positive Impact of Mobile Electronic Devices in Facilitating Activation in Information Seeking (+++)

- "MED certainly facilitate the interest in seeking information on health and medicines..." [student]
- "It would be definitely helpful to use MED for health-related information." [health professional]
- "They [MED] have an impact because everything that interests you can be found there, medicines, food, whatever you want." [patient]

Impact of Mobile Electronic Devices in the Patient-Physician Relationship (+++)

- "Now for the doctor, it has become a bit harder because patients now are [easily] informed about certain types of illnesses, and doctors need to be more careful not to go directly to the diagnosis and medication." [student]
- "When patients have access to this information, their rapport with doctors is changing a lot, and somehow it is thought that the doctor is replaceable because of technology..." [student]
- "If the patient gets that information from a site or trusted apps, it's pretty good that when you are in a 24-hour shift, you have 60 to 70 patients within a day, you cannot explain a lot to everyone, and if the patient would be appropriately informed it would be much better and easier." [health professional]

Potential Negative Impact of Mobile Electronic Devices in Patients (+++)

- "Inaccurate online information in patients who have no health information can lead to a worse condition due to the stress they create...We should not replace the doctor with the information that is on the Internet." [student]
- "The patients have never read more, but also, they have never read nonsense things more. It is good to read, but not so that in every portal is the fluid which heals the heart diseases, the fluid that heals the kidney disease..." [health professional]
- "If I used the internet, I would be schizophrenic. I do not object to the technique, but it is very wide." [patient]
- "The use of MED for information is a convenience, but it is not a security, so first we need to consult with the relevant doctor." [student]

The Role of Health Professionals in Facilitating the Use of MEDs to Enhance Patient Activation Toward Seeking Health-Related Information

Students considered that health professionals' role in providing information on health and medications is irreplaceable, and according to them, no matter what information can be obtained via the internet, the final source should be health professionals. According to students, there is currently a lack of time management by health professionals during patient contact, causing a lack of health information provision. Health professionals also see their role as a major responsibility but are nevertheless deficient because of the lack of available time. Patients also supported the importance and primary role of health professionals in providing information on health and medications. Regarding the role of health professionals in assisting with the use of MEDs, students suggested that patients should be referred to reliable databases or websites where they could be accurately informed, so the reading of unconfirmed scientific information would be prevented. Health professionals suggested that if they were more active in this role, it could have a significant impact, but this is not the case. Patients considered that if they were instructed to read about their health or medication via MEDs, they would do it and think it would positively affect their health. Subthemes with the comments of participants are shown in Textbox 5.

Textbox 5. Theme 5—summary of comments corresponding to each of the subthemes.

Health Professionals Have a Key Role in Providing Health Information (+++)

- "Despite receiving information through the internet, the main and accurate source should be the doctor." [student]
- "Informing the client about his/her condition for getting different medication should be done primarily by each health worker." [health professional]
- "Health professionals have the main role [ie, in providing health information]." [patient]

Health Professionals' Recommendation on Seeking Health-Related Information Through Mobile Electronic Devices (+++)

- "It would have been good to inform patients in this regard, but I think very few do. I think it would be much better if you visit a doctor and they help you with using an app that will help you in the future." [student]
- "I try to give the patient's family members information on a certain drug, and I recommend them to read on the Internet what we are prescribing." [health professional]
- "They accomplish this role very well; they tell us to read." [patient]

Deficient Role of Health Professionals in Regard to Assisting Patients With Mobile Electronic Devices Used for Health Information (+++)

- "They have a very important role, but I think they do not use it...they do not give the [exact] information, because of time or other engagements. I think it would have been good to inform patients in this regard [ie, using apps, reading health and medications information through MED], but I think very few do so." [student]
- "I do not believe it would be functional if a doctor recommends something to the patient to read..." [health professional]
- "It would have been very good if they would have assisted us in this aspect; I would have read and done so." [patient]

Discussion

Principal Findings

This study explored the impact of MEDs on activating patients to seek information on chronic conditions and medications. Data were collected from students, health professionals, and patients. The findings of this study suggest that MEDs could help activate patients to seek information on health and medications. MEDs provide quick access to the information required and provide a large amount of information, which could facilitate patient activation to seek health-related information. Furthermore, a previous observational study showed that mHealth apps were most useful for obtaining general health and physical activity and nutrition information, finding no significant difference in mHealth app' use on patients' socioeconomic status [21]. Therefore, the potential role of mHealth apps is empowering patients, especially those with lower socioeconomic status, to improve their health outcomes. Considering that there is evidence that supports the effectiveness of mHealth technologies in clinical outcomes and patient-centered care outcomes such as patient satisfaction and patient engagement [22], this study's findings are especially important as they further delineate the role of MEDs in facilitating patient activation toward seeking health-related information.

The findings of this study suggest a mutual acceptance of MEDs in patient activation to seek health-related information from patients, health professionals, and future health professionals. However, there were different reasons and approaches for all three subsamples when using the MEDs. Students' motives for seeking health and medication information through MEDs included easy and quick access and the need for constant information updates that assist them in their studies. Technological barriers were identified and were consistent with previous studies [23], suggesting the need for improvement in access to library-licensed mobile resources. Health professionals seek information through MEDs mainly on new drugs while maintaining a preference for books when searching for information on diseases. This is particularly the case with brand drug names, which may be unknown to some health care staff. In addition, the findings of this study suggest that MEDs in general and mobile apps, in particular, may have a positive impact on the work routine of professional health care staff as they enable them to make quicker decisions with a lower degree of error, thus improving practice efficiency and management of clinical cases. These results are consistent with the findings of previous studies [24,25]. In addition to requiring information on their medications, patients were also interested in knowing more about complementary medications, nutritionally advantageous diets, and healthy lifestyles in general. This is because MEDs provide access to vast amounts of information and a source of unlimited information. Therefore, although findings from all FGs suggest a potential positive role of MEDs in patient activation in seeking health-related information, there were different motives and barriers for each of the study subsamples.

Patients using MEDs and receiving information on their health conditions felt more empowered in managing their health and suggested that higher activation levels could be achieved via the use of MEDs. In contrast, patients who did not use MEDs indicated a lower level of activation and difficulty in understanding the actions required to maintain a healthy lifestyle. However, it was interesting to note that none of the patients identified the existence of nonreliable health-related web-based information as an issue when using MEDs. Therefore, it enhances the need for health professionals to discuss the validity of web-based information that patients may consult and apply to their health.

In addition, although patients are getting more involved in their health care decisions, the proportion who are willing to take

serious action and change their health behaviors is still in the minority. It is thus helpful to know which types of patients might become more actively engaged and apt to take such behaviors, and therefore the mechanisms that might be used to aid them and other patients [26]. Although MEDs could facilitate patient activation in seeking health-related information for some patients, the role of MEDs might be vague for some others. Therefore, besides evaluating the patient activation level, assessing the stage of change in the transtheoretical model (TTM) might help determine for which patients MEDs could affect patient activation in seeking health-related information.

Additional research might consider using the TTM, an integrative biopsychosocial model aimed at predicting one's likelihood of behavior change [27]. Using the TTM can help discern where patients are in their contemplation to engage and use mobile devices and technology for their health, and thus might be used in concert with the results of our study to design educational interventions and communication with patients.

This study suggests that health professionals can influence the activation of patients to seek health-related information using MEDs by assisting them in the use of MEDs to search for information, which would be an additional motive for patients. This guidance and assistance could occur during regular check-ups that patients with chronic conditions have. During this time, health care professionals could advise using certain credible websites to require health-related information and even try to conduct a web search related to the patient's specific concerns. This assistance would facilitate patients' interest in using MEDs for seeking information on health and medications. However, the health professionals' use of MEDs might influence their ability to support and guide patients in their use related to seeking health-related information. Evidence indicates that adequate training of health professionals on this matter would influence health education and improve the population's overall health on primary and secondary prevention bases [28]. However, it is promising that future health professionals tend to highly use MEDs, positing them in a better empowerment position to assist and advise their patients regarding seeking health-related information on the web than older cohorts of health professionals.

Furthermore, it is interesting to note that even patients who did not use MEDs for information on health and medications suggested that they would do so if health professionals would recommend specific information websites or mobile apps. The reliability of web-based health information remains a concern [29], and the quality of health information accessed by patients remains unevaluated [30]. Thus, as students and health professionals would recognize a valid data source, it would be questionable for patients with different training and backgrounds. In addition, evidence indicates that age differences play a vital role in credibility judgments among patients seeking health information on the web, showing that older adults have a higher tendency to passively accept web-based information compared with younger adults [31], thus suggesting a need for different training approaches to these populations. Finally, research indicates that the only predictor of mHealth use for self-management was patient information technology skills [32]. Therefore, health care professionals should advise all patients about MEDs, regardless of their age.

Study Limitations

This study had several limitations. First, the study lacks wider representativeness, as it was conducted in one city of Kosovo, and it did not include a wider range of participants who could be potential users of MEDs. This limitation can be considered minimal in the results obtained because participants originated from various parts of Kosovo, and a saturation point in terms of themes and subthemes was achieved even across diverse groups of individuals. The generalizability of this study's results is questionable because of the sample size and sample characteristics. However, considering that in 2019 the global internet access rate was 51.4%, and it was estimated that 86.7% of the population in developed countries had internet access [33], data from a study in 2017 showed that 88.5% of households in Kosovo had internet access [34]. Therefore, this study's findings would be more applicable in countries with similar internet access coverage. However, the qualitative *research* goal is not the generalizability as much as it is the generation of rich and contextualized understanding of unexplored phenomena [35].

In addition, FGs with health professionals consisted of specialists in various fields of medicine and nurses. The diversity of health professionals in FGs may have facilitated exploring different perspectives, although segmentation of these FG participants could have facilitated comparative data analysis. However, it has been previously reported that homogeneous groups of participants in FG meetings can capitalize on the shared experiences of participants [36]; thus, this approach was used to conduct separate FGs with students, health professionals, and patients. Finally, the FGs were conducted with participants in Kosovo, who were almost entirely of ethnic Albanian descent. As qualitative research, there was no instrumentation to translate directly; however, the concepts and theories from which the FG guide was constructed had their basis in the English language literature. Furthermore, this potential limitation should also be considered in lieu of the fact that regardless of location and language used, access to the internet, MEDs, and mobile apps has increased significantly worldwide; therefore, patients and health professionals are expected to exhibit similar behaviors when adopting technology. In this regard, it may be worth emphasizing that Kosovo is known to have the highest levels of household internet access in the world [37]. Nonetheless, we believe there is a unique internet- and media-related characteristic of our sample that may affect the use of MEDs to navigate health-related information. This uniqueness is derived from the fact that the Kosovo population has strong family and sociocultural ties with its large diaspora living overseas and with whom there is a high reliance on MEDs to exchange information on a regular basis. In this environment, MEDs users in Kosovo would be exposed to cross-cultural experiences derived from various societies that the Kosovo diaspora has influenced globally, which in turn may also have an impact on how they assess and interpret information. This characteristic of our sample as well as of similar populations groups merits further research to better understand the

implications on navigation via MEDs and the use of health-related information.

This study provides data referring to some of the basic motives for using MEDs for information search on health and medications and suggests the need for a change in health professionals' approach to assist using MEDs to facilitate patient activation in this regard. There is a need for further research into the clinical and economic impact of using MEDs in facilitating patients' activation to seek information on their chronic conditions and medications.

Conclusions

This study suggests that MEDs might help facilitate patients' activation to seek information on chronic conditions and medications. The motives for searching for information through MEDs related to the activation of patients have been identified. The findings suggest that health professionals' roles should be reconsidered to include additional assistance to their patients in using MEDs, specifically in recommending valid and trustworthy websites or mobile apps to search for information on chronic conditions and medications.

Acknowledgments

The authors wish to acknowledge and thank all of the FG participants of this study.

Conflicts of Interest

None declared.

References

1. Hibbard J, Gilburt H. Supporting people to manage their health: An introduction to patient activation. Kings Fund. 2014. URL: https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/supporting-people-manage-health-patient-activation-may14.pdf [accessed 2021-07-31]
2. Charmel PA, Frampton SB. Building the business case for patient-centered care. Healthcare Financial Management Association. 2008 Mar. URL: http://pqcnc-documents.s3.amazonaws.com/fpe/fperesources/BuildingBusinessCasePCCCharmelFrampton2008.pdf [accessed 2021-07-31]
3. Kilo CM, Wasson JH. Practice redesign and the patient-centered medical home: history, promises, and challenges. Health Aff (Millwood) 2010 May;29(5):773-778. [doi: 10.1377/hlthaff.2010.0012] [Medline: 20439860]
4. Greene J, Hibbard JH. Why does patient activation matter? An examination of the relationships between patient activation and health-related outcomes. J Gen Intern Med 2012 May;27(5):520-526 [FREE Full text] [doi: 10.1007/s11606-011-1931-2] [Medline: 22127797]
5. Hibbard JH, Greene J, Overton V. Patients with lower activation associated with higher costs; delivery systems should know their patients' 'scores'. Health Aff (Millwood) 2013 Feb;32(2):216-222 [FREE Full text] [doi: 10.1377/hlthaff.2012.1064] [Medline: 23381513]
6. Smith SG, Pandit A, Rush SR, Wolf MS, Simon C. The association between patient activation and accessing online health information: results from a national survey of US adults. Health Expect 2014 Dec 05;18(6):3262-3273. [doi: 10.1111/hex.12316]
7. Hibbard JH, Cunningham PJ. How engaged are consumers in their health and health care, and why does it matter? Res Brief 2008 Oct(8):1-9. [Medline: 18946947]
8. 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]
9. Siau K, Shen Z. Mobile healthcare informatics. Med Inform Internet Med 2006 Jun 12;31(2):89-99. [doi: 10.1080/14639230500095651] [Medline: 16777784]
10. O'Leary KJ, Lohman ME, Culver E, Killarney A, Randy SG, Liebovitz DM. The effect of tablet computers with a mobile patient portal application on hospitalized patients' knowledge and activation. J Am Med Inform Assoc 2015 Jun 15:159-165. [doi: 10.1093/jamia/ocv058] [Medline: 26078412]
11. Choi A, Lovett AW, Kang J, Lee K, Choi L. Mobile applications to improve medication adherence: existing apps, quality of life and future directions. Adv Pharmacol Pharm 2015 Sep;3(3):64-74. [doi: 10.13189/app.2015.030302]
12. Moura J, Almeida AM, Roque F, Figueiras A, Herdeiro MT. A mobile app to support clinical diagnosis of upper respiratory problems (eHealthResp): co-design approach. J Med Internet Res 2021 Jan 28;23(1):e19194 [FREE Full text] [doi: 10.2196/19194] [Medline: 33507153]
13. Pecina JL, Wyatt KD, Comfere NI, Bernard ME, North F. Uses of mobile device digital photography of dermatologic conditions in primary care. JMIR Mhealth Uhealth 2017 Nov 08;5(11):e165 [FREE Full text] [doi: 10.2196/mhealth.8257] [Medline: 29117934]
14. Martinez-Millana A, Jarones E, Fernandez-Llatas C, Hartvigsen G, Traver V. App features for type 1 diabetes support and patient empowerment: systematic literature review and benchmark comparison. JMIR Mhealth Uhealth 2018 Nov 21;6(11):e12237 [FREE Full text] [doi: 10.2196/12237] [Medline: 30463839]

15. Kaboli PJ, Howren MB, Ishani A, Carter B, Christensen AJ, Vander Weg MW. Efficacy of patient activation interventions with or without financial incentives to promote prescribing of thiazides and hypertension control: a randomized clinical trial. JAMA Netw Open 2018 Dec 07;1(8):e185017 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.5017] [Medline: 30646291]
16. Silver L, Smith A, Johnson C, Jiang J, Anderson M, Rainie L. Mobile connectivity in emerging economies. Pew Research Center. 2019. URL: https://www.pewresearch.org/internet/2019/03/07/mobile-connectivity-in-emerging-economies/ [accessed 2021-07-31]
17. Morgan DL. Focus groups. Annu Rev Sociol 1996 Aug;22(1):129-152. [doi: 10.1146/annurev.soc.22.1.129]
18. Stewart D, Shamdasani P. Focus Groups Theory and Practice. London: SAGE Publishing; 2014:1-224.
19. Williams M, Moser T. The art of coding and thematic exploration in qualitative research. International Management Review. 2019. URL: http://www.imrjournal.org/uploads/1/4/2/8/14286482/imr-v15n1art4.pdf [accessed 2021-08-03]
20. Hoti K, Hughes J, Sunderland B. Medication supply to residential aged care facilities in Western Australia using a centralized medication chart to replace prescriptions. BMC Geriatr 2012 Jul 11;12(1):25-26 [FREE Full text] [doi: 10.1186/1471-2318-12-25] [Medline: 22676640]
21. Ramirez V, Johnson E, Gonzalez C, Ramirez V, Rubino B, Rossetti G. Assessing the use of mobile health technology by patients: an observational study in primary care clinics. JMIR Mhealth Uhealth 2016 Apr 19;4(2):e41 [FREE Full text] [doi: 10.2196/mhealth.4928] [Medline: 27095507]
22. Bruce CR, Harrison P, Nisar T, Giammattei C, Tan NM, Bliven C, et al. Assessing the impact of patient-facing mobile health technology on patient outcomes: retrospective observational Cohort study. JMIR Mhealth Uhealth 2020 Jun 26;8(6):e19333 [FREE Full text] [doi: 10.2196/19333] [Medline: 32589161]
23. Boruff JT, Storie D. Mobile devices in medicine: a survey of how medical students, residents, and faculty use smartphones and other mobile devices to find information. J Med Libr Assoc 2014 Jan;102(1):22-30 [FREE Full text] [doi: 10.3163/1536-5050.102.1.006] [Medline: 24415916]
24. Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. Pharm Ther 2014 May;39(5):356-364 [FREE Full text] [Medline: 24883008]
25. Siebert JN, Lacroix L, Cantais A, Manzano S, Ehrler F. The impact of a tablet app on adherence to american heart association guidelines during simulated pediatric cardiopulmonary resuscitation: randomized controlled trial. J Med Internet Res 2020 May 27;22(5):e17792 [FREE Full text] [doi: 10.2196/17792] [Medline: 32292179]
26. Prochaska JO. Decision making in the transtheoretical model of behavior change. Med Decis Making 2008 Nov;28(6):845-849. [doi: 10.1177/0272989X08327068] [Medline: 19015286]
27. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997 Sep;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
28. Bert F, Giacometti M, Gualano MR, Siliquini R. Smartphones and health promotion: a review of the evidence. J Med Syst 2014 Jan;38(1):9995. [doi: 10.1007/s10916-013-9995-7] [Medline: 24346929]
29. Epstein HB. Cyberchondriacs. J Hosp Librariansh 2017 Oct 23;17(4):317-322. [doi: 10.1080/15323269.2017.1367901]
30. Diaz JA, Griffith RA, Ng JJ, Reinert SE, Friedmann PD, Moulton AW. Patients' use of the internet for medical information. J Gen Intern Med 2002 Mar;17(3):180-185 [FREE Full text] [doi: 10.1046/j.1525-1497.2002.10603.x] [Medline: 11929503]
31. Liao QV, Fu W. Age differences in credibility judgments of online health information. ACM Trans Comput-Hum Interact 2014 Feb;21(1):1-23. [doi: 10.1145/2534410]
32. Jiang J, Zhu Q, Zheng Y, Zhu Y, Li Y, Huo Y. Perceptions and acceptance of mHealth in patients with cardiovascular diseases: a cross-sectional study. JMIR Mhealth Uhealth 2019 Feb 04;7(2):e10117 [FREE Full text] [doi: 10.2196/10117] [Medline: 30714942]
33. Global internet penetration 2020. Statista. 2020 Nov. URL: https://www.statista.com/statistics/209096/share-of-internet-users-in-the-total-world-population-since-2006/ [accessed 2021-02-16]
34. Results of the survey on use of information and communication technology. Kosovo Agency of Statistics - Series 5: Social Statistics. 2017. URL: https://ask.rks-gov.net/media/3831/tik-2017-ang.pdf [accessed 2021-02-16]
35. Polit DF, Beck CT. Generalization in quantitative and qualitative research: myths and strategies. Int J Nurs Stud 2010 Nov;47(11):1451-1458. [doi: 10.1016/j.ijnurstu.2010.06.004] [Medline: 20598692]
36. Kitzinger J. Qualitative research. Introducing focus groups. Br Med J 1995 Jul 29;311(7000):299-302 [FREE Full text] [doi: 10.1136/bmj.311.7000.299] [Medline: 7633241]
37. Digital economy and society statistics - households and individuals. Eurostat. 2020. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php/Digital_economy_and_society_statistics_-_households_and_individuals [accessed 2021-03-02]

Abbreviations

FG: focus group
MED: mobile electronic device
mHealth: mobile health
TTM: transtheoretical model

UCCK: University Clinical Center of Kosovo

Edited by L Buis; submitted 06.12.20; peer-reviewed by LY Chen, D Zhang, M Harris; comments to author 11.01.21; revised version received 02.03.21; accepted 07.05.21; published 11.08.21.

Please cite as:
Imeri H, Desselle S, Hetemi D, Hoti K
Mobile Electronic Devices as Means of Facilitating Patient Activation and Health Professional Empowerment Related to Information Seeking on Chronic Conditions and Medications: Qualitative Study
JMIR Mhealth Uhealth 2021;9(8):e26300
URL: https://mhealth.jmir.org/2021/8/e26300
doi:10.2196/26300
PMID:34383686

Original Paper

Supervised Exercise Therapy Using Mobile Health Technology in Patients With Peripheral Arterial Disease: Pilot Randomized Controlled Trial

Katrin Paldán[1,2], Dr; Martin Steinmetz[3], Dr; Jan Simanovski[3], BA; Christos Rammos[3], Prof Dr; Greta Ullrich[3]; Rolf Alexander Jánosi[3], PD Dr; Susanne Moebus[1], Prof Dr; Tienush Rassaf[3], Prof Dr; Julia Lortz[3], PD Dr

[1]Institute for Urban Public Health, University Clinic of Essen, University of Duisburg-Essen, Germany, Essen, Germany

[2]Personal Analytics Centre of Competence, Department of Engineering Sciences, University of Duisburg-Essen, Essen, Germany

[3]Department of Cardiology and Vascular Medicine, West-German Heart and Vascular Center Essen, University of Duisburg-Essen, Essen, Germany

Corresponding Author:
Julia Lortz, PD Dr
Department of Cardiology and Vascular Medicine
West-German Heart and Vascular Center Essen
University of Duisburg-Essen
Hufelandstr 55
Essen, 45147
Germany
Phone: 49 201723 ext 84995
Email: julia.lortz@uk-essen.de

Abstract

Background: Mobile health interventions are intended to support complex health care needs in chronic diseases digitally, but they are mainly targeted at general health improvement and neglect disease-specific requirements. Therefore, we designed TrackPAD, a smartphone app to support supervised exercise training in patients with peripheral arterial disease.

Objective: This pilot study aimed to evaluate changes in the 6-minute walking distance (meters) as a primary outcome measure. The secondary outcome measures included changes in physical activity and assessing the patients' peripheral arterial disease–related quality of life.

Methods: This was a pilot two-arm, single-blinded, randomized controlled trial. Patients with symptomatic PAD (Fontaine stage IIa/b) and access to smartphones were eligible. Eligible participants were randomly assigned to the study, with the control group stratified by the distance covered in the 6-minute walking test using the TENALEA software. Participants randomized to the intervention group received usual care and the mobile intervention (TrackPAD) for the follow-up period of 3 months, whereas participants randomized to the control group received routine care only. TrackPAD records the frequency and duration of training sessions and pain levels using manual user input. Clinical outcome data were collected at the baseline and after 3 months via validated tools (the 6-minute walk test and self-reported quality of life). The usability and quality of the app were determined using the Mobile Application Rating Scale user version.

Results: The intervention group (n=19) increased their mean 6-minute walking distance (83 meters, SD 72.2), while the control group (n=20) decreased their mean distance after 3 months of follow-up (–38.8 meters, SD 53.7; *P*=.01). The peripheral arterial disease–related quality of life increased significantly in terms of "symptom perception" and "limitations in physical functioning." Users' feedback showed increased motivation and a changed attitude toward performing supervised exercise training.

Conclusions: Besides the rating providing a valuable support tool for the user group, the mobile intervention TrackPAD was linked to a change in prognosis-relevant outcome measures combined with enhanced coping with the disease. The influence of mobile interventions on long-term prognosis must be evaluated in the future.

Trial Registration: ClinicalTrials.gov NCT04947228; https://clinicaltrials.gov/ct2/show/NCT04947228

(JMIR Mhealth Uhealth 2021;9(8):e24214) doi:10.2196/24214

KEYWORDS

peripheral arterial disease; mHealth; supervised exercise therapy; motivation; artery; exercise; mobile applications; lifestyle; well-being

Introduction

Background

The circulatory disorders of peripheral arteries due to atherosclerotic lesions, also known as peripheral arterial diseases (PAD), are the third most frequent manifestations of cardiovascular diseases (CVD) after coronary and cerebrovascular arterial diseases [1].

A primary goal in CVD treatment is to slow down disease progression and avoid major adverse cardiac or limb events. Nonetheless, patients with PAD lag behind those with coronary artery disease (CAD) in terms of optimal treatment patterns [2-4]. Although the survival rates for CAD and PAD have improved worldwide, PAD still comes with a high individual burden regarding the quality of life (QoL) and associated disabilities [2].

The individual restrictions in the daily life of patients with PAD are more important than statistical facts regarding mortality and morbidity. Intermittent claudication causes a progressive reduction of the pain-free walking distance (PWD), and it is an expression of worsening PAD [4]. This decrease in physical capability results in declining mental health and reduces patients' QoL [5].

Supervised exercise therapy (SET) is a cornerstone in the conservative management of intermittent claudication [4], and it extends the PWD. Even though SET is easy to practice and highly cost-effective, adherence to regular SET performance is relatively low [6,7]. The underuse of exercise can be partly explained by the lack of institutional resources [8] and both patients' and physicians' lack of interest in exercise [4,9].

Mobile health (mHealth) technologies increase incentives and provide digital support for patients with PAD on several treatment levels [10-12]. They potentially lead to higher exercise training adherence and widen the scope of patient-centered health care [13], but so far, studies show opposite results [11,14]. While patients with PAD highly desire specific support tools, and app stores are inundated with health and fitness apps, PAD-specific solutions are presently lacking [15].

Objective

We developed a smartphone app named TrackPAD [16] to provide PAD-specific support for SET. This pilot study aims to evaluate the TrackPAD application as to its suitability and feasibility in outcome measures relevant to the prognosis of PAD by assessing the participant's 6-minute walking distance (meters).

Methods

Study Aims, Research Questions, and Outcomes

The TrackPAD pilot study aimed to answer the following research questions:

1. Is it feasible to implement the app into everyday practice?
2. Is TrackPAD suitable for recording patients' daily and weekly SET performance?
3. Does the TrackPAD improve the prognosis of PAD and related QoL?

The primary outcome was defined as the change in the 6-minute walking distance using a standardized protocol at baseline and after 3 months of follow-up [17]. The 6-minute walk test was performed under the supervision of a trained exercise technician. Participants were instructed to cover as much distance as possible, walking up and down a 50-meter hallway for up to 6 minutes. Participants were asked to push a measuring wheel during the entire 6 minutes of the test, but they could take breaks if necessary. They were also allowed to use an assistive device during both walking tests if they so desired. The technician stood in the middle of the course and supervised the walking test, but they did not encourage participants. The total distance walked in the test was read off of the measuring wheel. In patients with heart failure and reduced ejection fraction, a decreased 6-minute walking distance was associated with increased mortality, nonfatal cardiovascular events, and heart failure–related hospitalizations [18-20]. A decreased 6-minute walking distance was associated with a predictive value of mortality in patients with chronic obstructive pulmonary disease [21]. Among patients with PAD, the baseline 6-minute walking distance predicts rates for all-cause mortality, CVD mortality, and mobility loss [22,23]. Additionally, an incline of 20 meters was linked to a considerable improvement in total walking ability [24].

Aside from the 6-minute walking distance being objective and well-validated with respect to walking ability predicting mobility loss and mortality in PAD, it has an excellent test-retest reliability [25,26]. The 6-minute walking test offers several advantages over treadmill testing in PAD as it correlates more closely with physical activity levels and is not associated with the learning effect of performing repeated tests [27].

The secondary outcome measures were changes in physical activity and assessing the patient's PAD-related QoL via PAD-QoL. The PAD-QoL questionnaire is a validated PAD-specific questionnaire [28] containing five factors: (1) social relationships and interactions, (2) self-concept and feelings, (3) symptoms and limitations in physical functioning, (4) fear and uncertainty, and (5) positive adaptation. In addition, individual factors regarding sexual function, intimate relationships, and job function will also be assessed. An evaluation of the use of the TrackPAD app was also performed for the intervention group using the user version of the Mobile Application Rating Scale questionnaire. It provides a 20-item measure including 4 objective quality subscales for engagement, functionality, aesthetics, and information quality [29].

Study Design, Population, and the TrackPAD App

Study Design and Recruitment

This paper reports the results from the pilot study, including TrackPAD app usability tests for the target group (patients with PAD). In preparation for the pilot study, we conducted a recently published questionnaire study [15] evaluating the needs and requirements of designing mobile interventions for patients with PAD.

The TrackPAD pilot study was designed as a 2-armed randomized controlled trial and included patients with diagnosed and symptomatic PAD. It is a closed parallel-group trial (control and intervention groups were assessed simultaneously), with blinded assessors and face-to-face assessment components and a 3-month follow-up. Besides information regarding the pilot study, a call for participation was announced in a local newspaper (Westdeutsche Allgemeine Zeitung, local section for Essen and Duisburg) with the contact information provided, including the phone number and email address (trackPAD@uk-essen.de). In addition, potential participants were actively solicited during their outpatient clinic visits or their inpatient stay at the Department of Cardiology and Vascular Medicine, University Clinic of Essen. Willing patients were asked to register for the pilot trial at the front desk of the outpatient clinic.

Randomization

After screening based on inclusion and exclusion criteria and obtaining written informed consent, participants were randomized into 2 groups by the Center for Clinical Studies in Essen using the TENALEA software. The control group included participants with standard care and no further mobile intervention. The intervention group included participants receiving standard care and additional mHealth-based self-tracking of their physical activity using TrackPAD. The participants were stratified based on their 6-minute walking test (distances less than 362 meters, between 362 and 430 meters, and more than 430 meters) to ensure an even distribution of the walking speed between the two groups. After the randomization process, participants were not replaced, regardless of the reason for exclusion.

Both groups were strongly advised to continue with their SET according to the current standard guidelines [4]. Participants of the intervention group received additional access to the TrackPAD app, which complemented the patients' current treatment. The TrackPAD app was freely accessible to the intervention group. Besides the support provided during the installation procedure, the app did not require further technical maintenance. The only external contact during the follow-up occurred if participants requested technical support. A nonphysician member of the study team helped participants.

The baseline and follow-up examinations took place at the Department of Cardiology and Vascular Medicine outpatient clinic. They included a 6-minute walking test and a measurement of the ankle-brachial index (ABI). The ABI Measurements were conducted using a Doppler probe on the tibial and anterior artery locations. According to the current European Society of Cardiology (ESC) guideline, the highest value was used for calculation and divided by the highest systolic brachial Doppler pressure [4].

The patients were asked to fill out a questionnaire package at both time points, including self-reported physical activity, demographic characteristics, and the PAD-QoL questionnaire. The PAD-QoL was translated into German by a native speaker and was pretested on 5 PAD patients not included in the study sample. The pretest did not reveal the need for any adjustments.

Inclusion Criteria

Main inclusion criteria were diagnosed and symptomatic PAD of the lower extremities, defined as Fontaine stage IIa or IIb. Fontaine stage IIa indicated intermittent claudication with a walking distance of more than 200 meters, whereas Fontaine stage IIb indicated intermittent claudication with a walking distance of fewer than 200 meters [4]. Additionally, patients must have a personal smartphone suitable for downloading and using the TrackPAD app (IOS version greater than 11.0 or Android version greater than 5.0). A detailed list of the inclusion and exclusion criteria is shown in Textbox 1.

Textbox 1. Inclusion and exclusion criteria. ABI: ankle-brachial index; PAD: peripheral arterial disease; NYHA: New York Heart Association; CCS: Canadian Cardiovascular Society.

Inclusion Criteria

- 18 years of age or older
- Diagnosis of lower extremity PAD based on either an ABI greater or equal to 0.9 in at least one leg, invasive or noninvasive imaging of stenotic lower extremity artery disease, or endovascular or surgical revascularization of lower extremity artery disease
- PAD Fontaine stage 11a/b
- Smartphone with the capacity to use TrackPAD (Android version greater than 5.0 or IOS version greater than 11.0)
- Written informed consent prior to any study procedures, including a specified follow-up evaluation
- Best-medical treatment in the last 2 months per standard guidelines

Exclusion Criteria

- Wheelchair-bound, use of walking aid, or walking impairment due to another cause than PAD
- Below or above-knee amputation
- Acute or critical limb ischemia
- PAD Fontaine Stage I, III, or IV
- No German knowledge
- Severe cognitive dysfunction
- Congestive heart failure with NYHA III-IV symptoms
- Active congestive heart failure requiring the initiation or up-titration of diuretic therapy
- Angina pectoris with CCS class 3 to 4 symptoms, myocardial infarction, or stroke in the last 3 months
- Active arrhythmia requiring the initiation or up-titration of anti-arrhythmic therapy
- Severe valve disease

TrackPAD App

The mobile intervention TrackPAD was designed by Rocket Apes GmbH. There were no associations between the authors and the developer. Moreover, TrackPAD was only designed for study purposes and not commercial use. We did not change any content during the study period, and all content was frozen during the trial. The only dynamic component was the leaderboard, which was adjusted based on the training sessions performed by the participants. The participants set their weekly goal of SET units at the beginning of each week. As recommended by the 2017 ESC guideline [4], each unit included 30 minutes of SET. If participants did not go through an entire unit at once, there were 3 different options: taking breaks, continuing the unit after recovery, or quitting prematurely. After completing each unit (fully completed 30 min or not), user feedback was requested (Figure 1; see Feedback after SET unit).

Figure 1. Main views of the TrackPAD-app.

To account for a PAD-tailored solution, we included the following features (Figure 1):

1. Weekly goal adaptation: The app suggested a new weekly goal using an internal algorithm based on the completion rate of a user's SET units during the previous week.
2. Feedback after SET units: The feedback after each SET unit contained PAD-specific information regarding leg pain levels, breathing, and overall exhaustion. Patients had to respond by choosing between 1 (minimum leg pain, no restriction in breathing, or minimum exhaustion) and 10 (maximum leg pain, maximum restriction in breathing, or maximum exhaustion) for each item.
3. Claudication reminder: Each SET unit started with a short reminder that the walking pace and incline must be adapted to reach a certain level of claudication to extend the PWD sustainably. The reminder popped up when each new SET unit was initiated and needed to be actively confirmed.
4. Personal achievements: The personal progress of each user was recorded to unlock achievement medals (eg, a notable increase in users' physical activity, activity performed during public holidays, or successes like an increase of performed SET units per week).
5. Leaderboard: The leaderboard contained different categories (ie, number of steps in single training sessions, number of completed training sessions, total minutes of physical activity, and percent increase of physical activity). The different leaderboards showed individual placements compared to other users using TrackPAD.
6. Patient events: Information on upcoming Department of Cardiology and Vascular Medicine patient events focusing on vascular diseases were stored and easily accessible via the main menu.
7. PAD-FAQ: An FAQ section was included to address common technical issues, important contact information, and general training advice. Instructions in case of increasing or new pain during the training were also included.

Ethics Approval and Consent to Participate

The local ethics committee of the University of Duisburg-Essen (18-8355-BO) approved this study. Written informed consent was collected from each participant before any study procedures, and contact information was delivered to each participant. Any changes will be communicated to the ethics committee. The pilot study started at the beginning of November 2018 and ended in March 2019.

Data Collection and Security

Data were stored on an encrypted European server. No personalized data were shared with the developer, and they were only accessible to the study team.

Sample Size Considerations and Statistical Analysis

To allow for missing data and loss to follow up, we aimed to recruit 23 to 25 participants per study arm. The results achieved an estimated power of t=0.46 (post hoc power analysis; Cohen d=0.5; P=.05; $F_{1,46}$=1.157). We used a two-tailed t-test, and the enrollment goal was 20 participants each for the intervention and control groups. P<.05 was estimated as the significance threshold. For sample size consideration and statistical analysis, we used R (version 3.6.0). To account for the heterogeneity of the walking distance to be covered, the analysis was performed separately for Fontaine stage IIa and IIb. The regression model was estimated by ordinary least squares and a differences-in-differences approach.

Results

Study Population and Baseline Characteristics

After screening and randomization, we included 46 participants in the pilot study, of whom 22 (48%) were randomized to the intervention group, and 24 (52%) were randomized to the control. During the follow-up, 7 (15%) participants dropped out, mainly due to personal reasons. For example, 5 (11%) participants withdrew due to the severe illness of a close relative, and 2 (4%) participants dropped out as a result of either worsening of a nonstudy-related disease or death (Figure 2; see Panel A). Table 1 shows a summary of the remaining participants' baseline characteristics.

Figure 2. Quantitative development of screened patients including reasons for dropouts and exclusions.

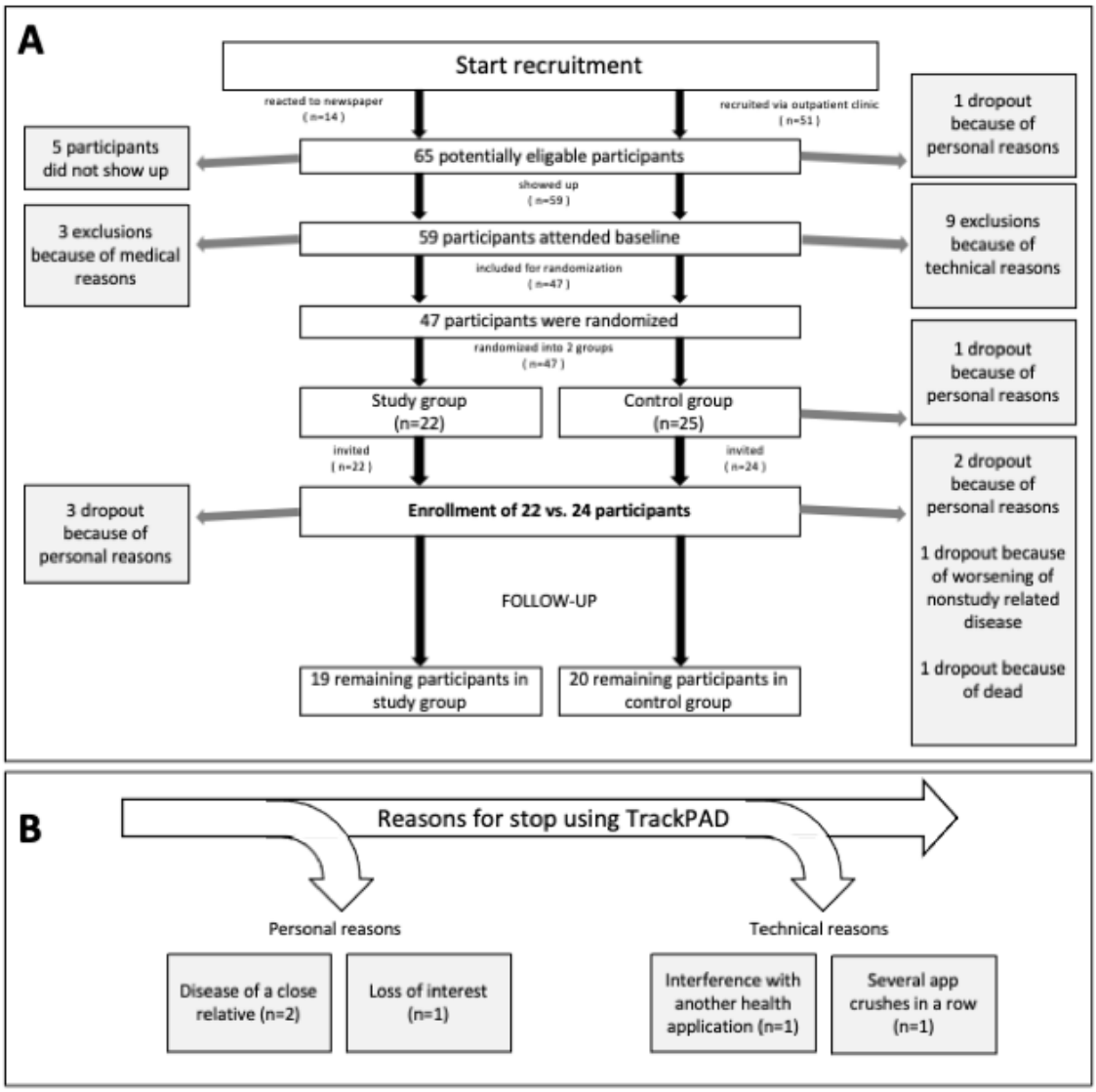


Table 1. Patient characteristics at baseline.

	Intervention group (n=19)	Control group (n=20)	*P* value
Age (years), mean (SD)	64.6 (9.8)	65.6 (7.7)	.72
Sex (male), n (%)	12 (63)	9 (45)	.34
Obesity (BMI > 30 kg/m^2), n (%)	5 (26)	1 (5)	.16
Prior MI[a], n (%)	2 (11)	3 (15)	.85
Hypertension, n (%)	12 (63)	16 (80)	.41
Diabetes, n (%)	4 (21)	6 (30)	.21
Hyperlipidemia, n (%)	12 (63)	13 (65)	.42
Previous peripheral intervention, n (%)	8 (42)	5 (25)	.26
Previous peripheral bypass graft, n (%)	3 (16)	5 (25)	.68
Previous PCI[b], n (%)	4 (21)	6 (30)	.69
Heart failure, n (%)	2 (11)	3 (15)	.85
Coronary arterial disease, n (%)	6 (32)	9 (45)	.51
Active/Former smoker, n (%)	6/11 (32/58)	8/10 (40/50)	.89
Fontaine stage IIa, n (%)	12 (63)	14 (70)	.44
Fontaine stage IIb, n (%)	7 (37)	6 (30)	.85
6-minutes walking distance (meters), mean (SD)	407 (80.8)	390.1 (66)	.35
ABI[c]	0.75 (0.21)	0.73 (0.18)	.46
Reported physical activity (days per week), mean (SD)	2.4 (1.4)	2.3 (1.9)	.35

[a]MI, myocardial injury.

[b]PCI, percutaneous coronary intervention.

[c]ABI, ankle-brachial index.

Increase in the 6-minute Walking Distance as a Primary Outcome

Of the 20 participants who increased their 6-minute walking distance at follow-up, 18 (90%) belonged to the intervention group using TrackPAD. The remaining participant in the intervention group did not change his covered distance at follow-up. In contrast, except for 2 (10%) participants, 18 (90%) participants in the control group showed decreased 6-minute walking distance at follow-up.

The mean distance covered in the 6-minute walking test showed a significant increase in the intervention group overall (83.0 meters, SD 72.2), whereas the mean walking distance of the control group decreased on average (–38.8 meters, SD 53.7; *P*<.001).

Both Fontaine stages showed similar trends, but the mean distance increase for the less progressed Fontaine stage IIa was more pronounced (intervention group: 97.0 meters, SD 78.6 vs. the control group: –35.3 meters, SD 55.9; *P*<.001). The Fontaine stage IIb showed a slight increase in mean walking distance for the intervention group (59.0 meters, SD 57.0) compared to the control group (–7.0 meters, SD 52.2), but it was still significant (*P*=.01).

TrackPAD was linked to a mean increase in the 6-minute walking distance of the intervention group, regardless of the Fontaine stage (95% CI 48.2-117.8). In contrast, the control showed either a slight or missing increase (95% CI –63.9-3.6). In total, the difference between both means was 121.8 meters (Fontaine stage IIa: 132.3 meters; IIb: 106.4 meters). Depending on the Fontaine stage, this resulted in a 17% (IIb) to 23% (IIa) increase of the covered distance at follow-up (Table 2).

Table 2. Differences in the 6-minute walking distance within and between study and control group after 3 months of follow-up.

	Fontaine IIa (n=26)		Fontaine IIb (n=13)		Fontaine IIa, IIb (n=39)	
	Study (n=12)	Control (n=14)	Study (n=7)	Control (n=6)	Study (n=19)	Control (n=20)
Difference in mean[a] (meters)	97.0	–35.3	59.4	–47.0	83.0	–38.8
Median (meters)	89.9	–22.0	30.0	–22.5	60	–22.0
SD (meters)	78.6	55.9	57.0	52.2	72.2	53.7
95% CI[b] (meters)	47.0-147.0	–67.5-3.0	6.3-111.8	–101.8-7.8	48.2-117.8	–63.9-3.6
Difference in mean between both groups (meters)	132.3		106.4		121.8	
SD (meters)	135.5		71.6		176.4	
95%-CI[c] (meters)	75.5-189.0		39.2-172.8		80.2-163.4	
P value	.01		.01		.01	

[a] Positive mean indicates an improvement.

[b]Difference between study and control group of the sub group.

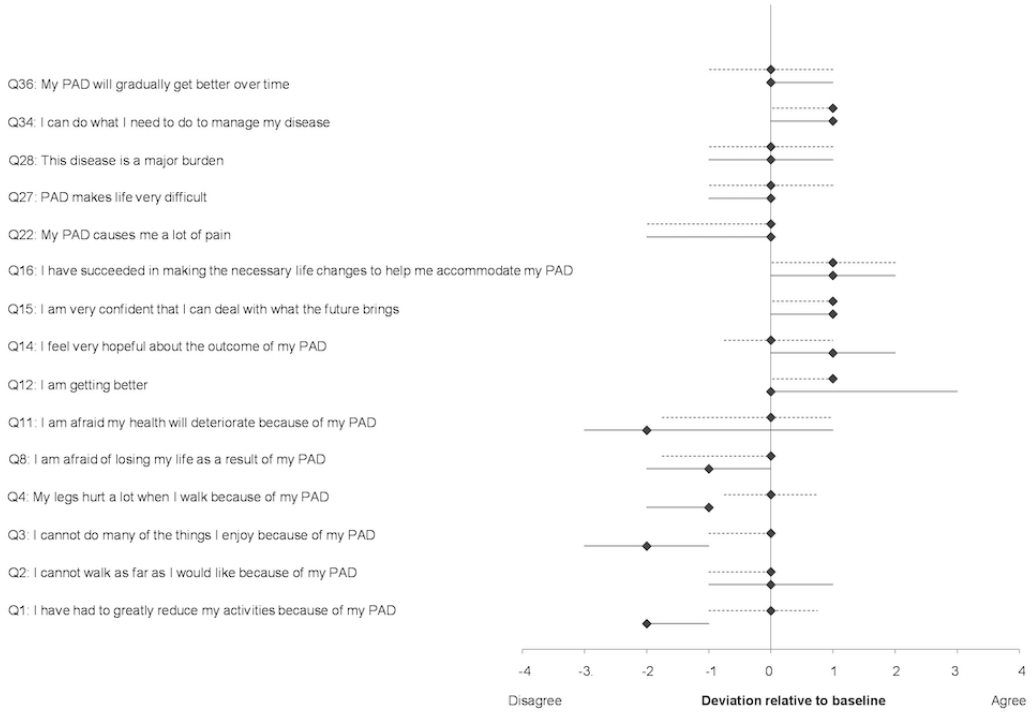
[c]The true difference of the population between both groups.

A difference-in-difference regression with fixed effects for time (accounting for a progression of PAD) and individual participant (accounting for unobserved heterogeneity between the participants) estimating the percentage change in the treatment effect showed that the effect of receiving access to TrackPAD increased the 6-minutes walking distance about 28% (SE 0.04). This effect was significant to a confidence level of 99%.

PAD-related Quality of Life

The PAD-related quality of life (PAD-QoL) was assessed by the PAD-QoL questionnaire at baseline and follow-up. No relevant differences were observed at baseline between both groups. However, at follow-up, significant changes were noted in 3 factors of the PAD-QOL, with the most extensive change evident in the "symptoms and limitations in physical functioning." The intervention group reported reduced limitations in their daily activity: "I have had to greatly reduce my activities because of my PAD" (Q1, intervention group: –1.6 meters, SD 1.4 vs control group: –0.1 meters, SD 1.0; *P*=.01); "I cannot do many of the things I enjoy because of my PAD" (Q3, intervention group: –1.8 meters, SD 1.5 vs control group: –0.4 meters, SD 1.0; *P*=.01); and "My legs hurt a lot when I walk because of my PAD" (Q4, intervention group: –1.4 meters, SD 1.3 vs control group: 0 meters, SD 1.1; *P*=.01). The intervention group also showed a change in a single item of the factor "fear and uncertainty," reporting a reduced fear of losing life because of PAD: "I am afraid of losing my life as a result of my PAD" (Q8, intervention group: –1.3 meters, SD 1.5 vs control group: –0.3 meters, SD 1.5; *P*=.048), and a change in the section "positive adaptation": "I feel very hopeful about the outcome of my PAD" (Q14, intervention group: 1.2 meters, SD 1.1 vs control group: 0.2 meters, SD 1.4; *P*=.02; Figure 3).

Figure 3. Excerpt of results from the PAD-QoL questionnaire survey [21]. Shown are the mean deviations relative to the baseline (diamond) and the 25th and 75th percentiles (lines). Study (solid line) and control group (dotted line) are plotted separately. PAD: peripheral arterial disease.



Overall, changes in the PAD-QoL over the 3 months of follow-up showed a less intense subjective symptom perception and fewer limitations in daily life among the intervention group.

Reported Physical Activity

To compare the two groups in terms of physical endurance at baseline, we recorded the reported physical activity. Both groups did not differ in days of physical activity per week (intervention group: 2.9 days per week, SD 2.8 vs control group: 2.4 days per week, SD 1.9; *P*=.44). Both groups had participants who were active for a median of 30 to 60 minutes (intervention group: n=9, 20% vs control group: n=4, 9%). In total, 12 participants (26%) were active for more than 60 minutes (intervention group: n=3, 7% vs control group: n=9, 20%). Among participants who exercised for less than 30 minutes weekly, 5 (11%) participants trained between 10 and 30 minutes weekly (intervention group: n=1, 2% vs control group: n=4, 9%), and 9 (20%) participants exercised less than 10 minutes weekly (intervention group: n=6, 13% vs control group: n=3, 7%).

At follow-up, 37 (80%) participants reported an increase in their weekly physical activity (intervention group: n=15, 33% vs control group: n=16, 35%), resulting in a comparable rise in physically active days per week in both groups (intervention group: plus 0.3 days per week, SD 3.5 vs control group: plus 0.4 days per week, SD 2.6; *P*=.93).

App Evaluation

App Usage

We considered intervention participants as active users if they performed at least 1 weekly training. During week 1, every participant was active. A dip from 19 (100%) to 14 (74%) active users was observed in week 2, increasing to 17 (89%) active users in week 3. During the following weeks, the activity remained stable, with 14 to 15 (70% to 75%, respectively) active users from week 5 to 12 (Table 3). During the 12 weeks of follow-up, the number of training sessions per week stayed roughly the same for the participants that remained active users.

Table 3. TrackPAD-app usage of the intervention group during the 12 weeks of follow-up.

Week	1	2	3	4	5	6	7	8	9	10	11	12
Total training sessions (units)[a]	66	48	53	74	60	61	49	57	58	51	55	48
Active user[b] (n)	19	14	17	16	15	15	15	14	15	14	14	15
Intervals per training session (units), mean[c] (SD)	2.5 (2.8)	2.8 (2.2)	2.6 (1.7)	2.0 (1.2)	1.8 (1.2)	2.3 (1.6)	2.7 (2.8)	2.4 (2.0)	2.2 (2.3)	2.1 (1.9)	2.5 (2.1)	2.7 (4.1)

[a]Total number of recorded training sessions for the respective week as assessed by the TrackPAD-App.

[b]Active user with at least one training interval in the corresponding week.

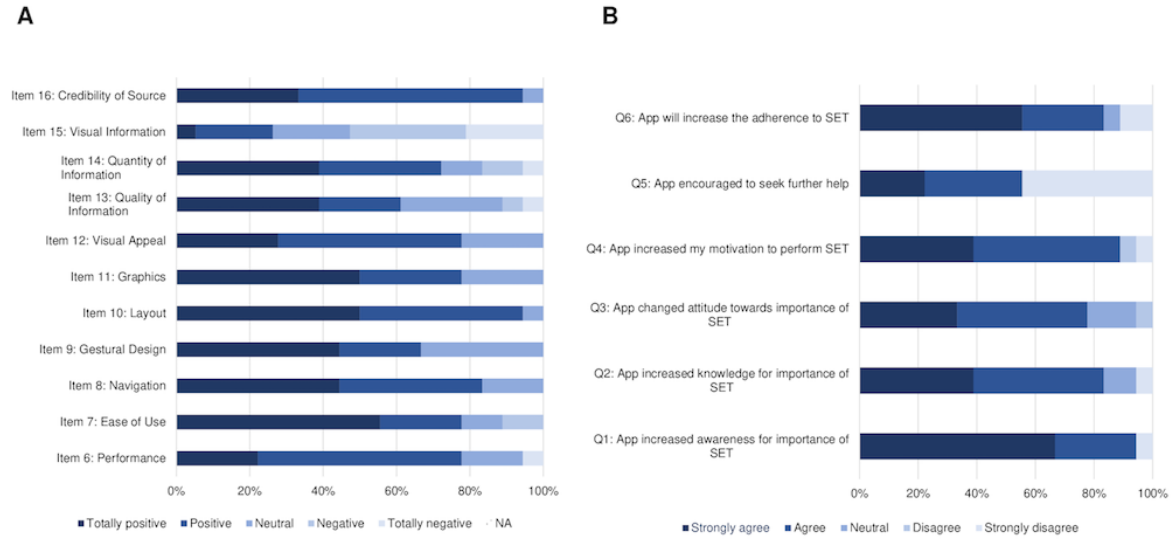
[c]Mean number of intervals during one training session. Each training session could be paused if necessary, resulting in each training session being subdivided into several intervals.

The reasons for discontinued TrackPAD use by the nonactive users (n=5, 26%) from week 5 onward were assessed at follow-up. Reasons for discontinued TrackPAD use were related to personal circumstances (n=3, 16%) and technical issues (n=2, 11%). Of the 2 participants who stopped using TrackPAD for personal reasons, 1 was due to the illness of a close relative, and the other lost interest. One participant stopped using TrackPAD due to reported interference between the TrackPAD app and their Samsung Health app, and another participant stopped the training sessions due to several sequential app crashes (Figure 2; see Panel B).

User Feedback

The vast number of questions regarding functionality, aesthetics, and informational content of TrackPAD were reported as positive to extremely positive (4 or 5 stars out of 5; Figure 4; see Panel A). However, the visual information provided within the app showed potential for improvement (Figure 4; see Panel A, Item 15); for example, the plausibility and correctness of descriptions represented by pictograms or pictures. Participants described this item mainly as “largely unclear.” Only 5 (25%) participants described the visual information as “mostly clear”(n=4) or “absolutely clear” (n=1).

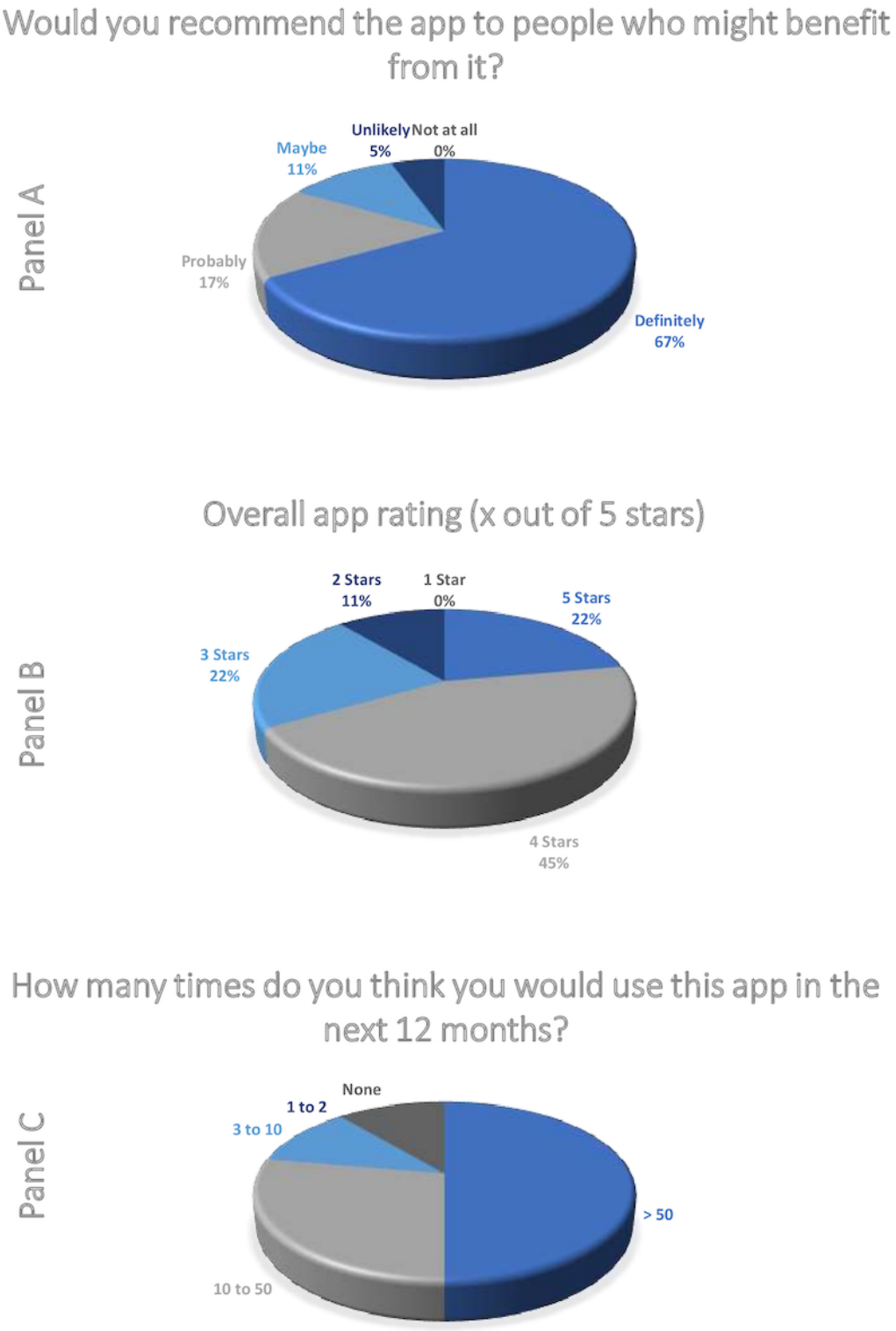
Figure 4. Participants' statements regarding the trackPAD app in terms of functionality, aesthetics and information according to the user version of the Mobile Application Rating Scale [22].



The users’ feedback also included questions regarding the perceived impact of the TrackPAD with respect to their PAD disease (Figure 4; see Panel B). Only 1 (6%) user disagreed, stating the app had not changed their awareness of SET (Q1). The other participants reported that the app had significantly increased their motivation to perform SET (Q4) and their compliance to SET (Q6). They also stated that using the app changed their attitude regarding SET (Q3) and increased their knowledge about SET (Q2).

Most users evaluated the app in all of the 3 categories positively. Only 3 (17%) users would “maybe” or are “unlikely” to recommend the app to people with existing PAD disease. Supporting the positive evaluation illustrate in Panel A (Figure 5), 13 (68%) users rated the app with at least 4 out of 5 stars (Figure 5; see Panel B). Future app use, at least every week, was reported by 10 (53%) users.

Figure 5. App rating of the study group after study end according to the user version of the Mobile Application Rating Scale [22].



The data underlying this article will be shared at reasonable request to the corresponding author.

Discussion

The implementation of novel technologies, specifically mobile interventions, can substantially change the landscape for the treatment of CVD [12,30]. General benefits of mHealth technologies include the wide reachability and the possibility of continuous access [31]. Although PAD represents a subgroup of CVD, patient characteristics and disease-specific requirements differ substantially from those patients with other CVD. Therefore, disease- and patient-tailored solutions are essential to the development of mobile interventions. One significant difference between the PAD population and patients with other CVD is the older age and the fact that the patient-centered development process needs to be expanded by one additional dimension. Previous studies already explored the use and

acceptability of mobile technologies in health care related to the users' age and identified age as an important factor in the design of mobile interventions, requiring greater technical support and reporting lower acceptability of using mobile technologies [32,33]. As such, there are measurable influences on intermediate outcomes (eg, increased satisfaction with care) and health outcomes (eg, better metabolic control) [15,34-36].

SET is one of the most relevant interventions in the conservative treatment of PAD, but barriers to exercise are still high. Besides low motivational aspects, intermittent claudication limits the sustainability of regular SET performance. Moreover, the requirements of primary care for patients with PAD focus on other priorities other than CVD in general [15,30]. To meet this specific patient population's needs and requirements, we proposed using a PAD-tailored mobile intervention to encourage the SET performance in patients with PAD.

Mobile technologies are increasingly used for health purposes, even among older adults who have demonstrated a lower uptake of technologies compared to younger people [37]. Although these technologies have the potential to assist in care coordination activities, like regular SET performance, most mobile apps are not designed specifically for this population which has complex health care needs and is older than the typical app user. The activity recognition mechanism of most mobile apps cannot accommodate the wide range of human movement linked to mobility impairment.

In this study, we gathered TrackPAD use input from the patients' perspective, and we observed a high level of user acceptance. Overall, we found satisfaction in terms of functionality, aesthetics, and informational content. Studies combining eHealth and PAD are rare, but the same trend of mobile technology user acceptance was observed in patients with noncommunicable diseases. A review of eHealth interventions for cancer survivors showed mobile interventions are promising tools [38]. Future work will need to examine the extent to which personalized activity recognition can support the diversity of movement.

Improvement was demonstrated through the visual information within TrackPAD and the clear assignment of pictograms or pictures. The weakness of the gestural concept resulted from the advanced age of the user group, which is often inexperienced in using mHealth and requires an age-adapted presentation [39]. Besides relevant barriers for older adults, lack of desire, costs, privacy and security considerations, visual acuity, and hand-eye coordination were important factors with respect to the acceptance of telehealth interventions [40]. These barriers will be adapted accordingly to improve the TrackPAD app following a patient-centered approach.

Since we designed a platform for both iOS and Android, some technical issues occurred due to the different technical implementations of the provider. The various mechanisms for counting steps presented a considerable challenge in designing a comparable app for both platforms. Depending on the manufacturer, step counts work either over a physical hardware mechanism and a software-based solution. This issue might become less relevant when it comes to personal use [41,42], but it also limits the analysis within a clinical trial. Because of the low number of smartphones that use the software-based solution, this issue did not occur in earlier tests. However, the disproportionate share of older mobile phones lacking a physical hardware mechanism within the intervention group revived the issue. In further trials, the inclusion of newer operating versions of Android mobile phones should be considered since this issue was only found in Android-based operating systems.

The disadvantages of simple activity tracking are known and common limitations in studies. The performance of systems trained with data in the laboratory setting substantially deteriorates when tested in real-life conditions [43]. Possible solutions might be user-calibration processes or the use of specified study-related devices to gain comparability.

Comparing the 6-minute walking distance between both groups in our study, we saw a significant increase in the mean walking distance of 80 meters in the intervention group using TrackPAD. Remarkably, we did not find any decrease in the walking distance within the intervention group, whereas 90% (n=18) of the control group did worse at follow-up compared to baseline. One reason for the longer walking distance might be because of the younger age of the study participants. Previous studies reported a mean age of more than 70 years, whereas the intervention group using TrackPAD had a mean age of 64. The higher increase may also be due to comparatively minor restrictions since two-thirds of the participants were classified as Fontaine stage IIa (PWD of more than 200 meters). The ease of initiating exercise among the Fontaine stage IIa patients with PDA compared to patients with higher Fontaine stages might be linked to better endurance during exercise and higher motivation in general. Moreover, the small sample size allows for substantial individual changes within the intervention group, leading to an upward deviation.

Although the covered distance in the 6-minute walk test only increased significantly in the intervention group, the self-reported physical activity increased in both groups at follow-up. An accurate assessment of physical activity using the PDA-QoL questionnaire seems questionable in the entire study population and has previously been described as a common issue [44,45]. Digital interventions also increase the potential to track background activity (ie, receiving objective statements in terms of physical activity) and will be considered in future trials. The recording of activity highs and lows throughout the day might also help identify optimal time points to send digital motivation notifications. It is important to note that messages can also decrease productivity if delivered at the wrong time points. Algorithms based on collected personalized information in smartphones might reduce the number of poorly-timed interruptions [46].

We also observed an increase in PAD-QoL regarding "symptoms and limitations in physical functioning" within the intervention group. The association between increased physical activity and an increased PAD-QoL has been reported in other studies [47-49]. In addition, the increased 6-minute walking distance was linked to better physical aspects of quality of life in participants with intermittent claudication, supporting its value as an outcome measure.

The main limitation of this study was the small sample size of the intervention group. Since we have analyzed some patient

characteristics (ie, Fontaine stage IIa and IIb) separately, the sample size per group decreased even further. However, the Fontaine stage allowed us to control for the differences in the participants' physical capability. Although we saw a relevant change in the primary outcome variable after follow-up, recordings of background activity during the follow-up period were available due to privacy restrictions. Based on the study design of this pilot, no blinding of the study participants was feasible, and motivational differences must be considered. Further research is needed to address this issue.

Using the smartphone–based tool TrackPAD, we found a significant increase in the mean 6-minute walking distance at follow-up, indicating a prognostically relevant change in walking ability in patients with moderate PAD. TrackPAD also bolstered a shift in the subjective symptom perception and fewer noticed limitations in terms of PAD-QoL. Thus, the TrackPAD app seems feasible and suitable for the target group of patients with PAD in terms of SET performance. Participants substantially valued the experience of using an app in the management of their care. Still, a further adaption of the visual presentation and the gestural concept that follows a patient-centered approach is needed.

Acknowledgments

This work was supported by the Stiftung Universitätsmedizin (D/106-21637) and the Deutsche Forschungsgesellschaft (DFG; DFG 969RA/12-1).

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 381 KB - mhealth_v9i8e24214_app1.pdf]

References

1. Song P, Rudan D, Zhu Y, Fowkes FJI, Rahimi K, Fowkes FGR, et al. Global, regional, and national prevalence and risk factors for peripheral artery disease in 2015: An updated systematic review and analysis. The Lancet Global Health 2019 Aug;7(8):e1020-e1030. [doi: 10.1016/s2214-109x(19)30255-4]
2. Bauersachs R, Zeymer U, Brière JB, Marre C, Bowrin K, Huelsebeck M. Burden of coronary artery disease and peripheral artery disease: A literature review. Cardiovasc Ther 2019;2019:8295054 [FREE Full text] [doi: 10.1155/2019/8295054] [Medline: 32099582]
3. Timmis A, Townsend N, Gale C, Torbica A, Lettino M, Petersen S. European Society of Cardiology: Cardiovascular disease statistics 2019. Eur Heart J. 2020 2020;41(1):12-85. [doi: 10.1093/ehjqcco/qcz065]
4. Victor A, Jean-Baptiste R, Marie-Louise E L B, Martin B, Marianne B, Tina C, et al. 2017 ESC Guidelines on the diagnosis and treatment of peripheral arterial diseases, in collaboration with the European Society for Vascular Surgery (ESVS). Russ J Cardiol 2018 Sep 09;ESC Guidelines on the DiagnosisTreatment of Peripheral Arterial Diseases, in collaboration with the European Society for Vascular Surgery (ESVS)(8):164-221. [doi: 10.15829/1560-4071-2018-8-164-221]
5. Bouwens E, van Lier F, Rouwet EV, Verhagen HJM, Stolker RJ, Hoeks SE. Type D personality and health-related quality of life in vascular surgery patients. Int.J. Behav. Med 2019 Jul 01;26(4):343-351. [doi: 10.1007/s12529-018-09762-3]
6. Harwood A, Smith GE, Cayton T, Broadbent E, Chetter IC. A systematic review of the uptake and adherence rates to supervised exercise programs in patients with intermittent claudication. Ann Vasc Surg 2016 Jul;34:280-289. [doi: 10.1016/j.avsg.2016.02.009] [Medline: 27126713]
7. Abaraogu UO, Abaraogu OD, Dall PM, Tew G, Stuart W, Brittenden J, et al. Exercise therapy in routine management of peripheral arterial disease and intermittent claudication: a scoping review. Therapeutic Advances in Cardiovascular Disease 2020 May 27;14:175394472092427. [doi: 10.1177/1753944720924270]
8. Rezvani F, Heider D, Härter M, König HH, Bienert F, Brinkmann J, et al. Telephone health coaching with exercise monitoring using wearable activity trackers (TeGeCoach) for improving walking impairment in peripheral artery disease: study protocol for a randomised controlled trial and economic evaluation. BMJ Open 2020 Jun 04;10(6):e032146 [FREE Full text] [doi: 10.1136/bmjopen-2019-032146] [Medline: 32503866]
9. Pedersen BK, Saltin B. Exercise as medicine: Evidence for prescribing exercise as therapy in 26 different chronic diseases. Scand J Med Sci Sports 2015 Dec 25;25 Suppl 3:1-72. [doi: 10.1111/sms.12581] [Medline: 26606383]
10. Gandapur Y, Kianoush S, Kelli HM, Misra S, Urrea B, Blaha MJ, et al. The role of mHealth for improving medication adherence in patients with cardiovascular disease: a systematic review. Eur Heart J Qual Care Clin Outcomes 2016 Oct 01;2(4):237-244 [FREE Full text] [doi: 10.1093/ehjqcco/qcw018] [Medline: 29474713]

11. Haveman ME, Kleiss SF, Ma KF, Vos CG, Ünlü Ç, Schuurmann RC, et al. Telemedicine in patients with peripheral arterial disease: Is it worth the effort? Expert Rev Med Devices 2019 Sep 05;16(9):777-786. [doi: 10.1080/17434440.2019.1649595] [Medline: 31379218]
12. Steinmetz M, Rammos C, Rassaf T, Lortz J. Digital interventions in the treatment of cardiovascular risk factors and atherosclerotic vascular disease. Int J Cardiol Heart Vasc 2020 Feb;26:100470 [FREE Full text] [doi: 10.1016/j.ijcha.2020.100470] [Medline: 32021904]
13. Argent R, Daly A, Caulfield B. Patient involvement with home-based exercise programs: Can connected health interventions influence adherence? JMIR Mhealth Uhealth 2018 Mar 01;6(3):e47 [FREE Full text] [doi: 10.2196/mhealth.8518] [Medline: 29496655]
14. McDermott M, Spring B, Berger J, Treat-Jacobson D, Conte M, Creager M. Effect of a home-based exercise intervention of wearable technology and telephone coaching on walking performance in peripheral artery disease: The HONOR randomized clinical trial. Journal of Vascular Surgery 2018 Sep;68(3):933. [doi: 10.1016/j.jvs.2018.07.014]
15. Lortz J, Simanovski J, Kuether T, Kreitschmann-Andermahr I, Ullrich G, Steinmetz M, et al. Needs and requirements in the designing of mobile interventions for patients with peripheral arterial disease: Questionnaire study. JMIR Form Res 2020 Aug 04;4(8):e15669 [FREE Full text] [doi: 10.2196/15669] [Medline: 32663154]
16. Paldán K, Simanovski J, Ullrich G, Steinmetz M, Rammos C, Jánosi RA, et al. Feasibility and clinical relevance of a mobile intervention using TrackPAD to support supervised exercise therapy in patients with peripheral arterial disease: study protocol for a randomized controlled pilot trial. JMIR Res Protoc 2019 Jun 26;8(6):e13651 [FREE Full text] [doi: 10.2196/13651] [Medline: 31244477]
17. ATS Committee on Proficiency Standards for Clinical Pulmonary Function Laboratories. ATS statement: guidelines for the six-minute walk test. Am J Respir Crit Care Med 2002 Jul 01;166(1):111-117. [doi: 10.1164/ajrccm.166.1.at1102] [Medline: 12091180]
18. Giannitsi S, Bougiakli M, Bechlioulis A, Kotsia A, Michalis LK, Naka KK. 6-minute walking test: a useful tool in the management of heart failure patients. Ther Adv Cardiovasc Dis 2019 Aug 23;13:1753944719870084 [FREE Full text] [doi: 10.1177/1753944719870084] [Medline: 31441375]
19. Keteyian SJ, Patel M, Kraus WE, Brawner CA, McConnell TR, Piña IL, HF-ACTION Investigators. Variables measured during cardiopulmonary exercise testing as predictors of mortality in chronic systolic heart failure. J Am Coll Cardiol 2016 Feb 23;67(7):780-789 [FREE Full text] [doi: 10.1016/j.jacc.2015.11.050] [Medline: 26892413]
20. Wegrzynowska-Teodorczyk K, Rudzinska E, Lazorczyk M, Nowakowska K, Banasiak W, Ponikowski P, et al. Distance covered during a six-minute walk test predicts long-term cardiovascular mortality and hospitalisation rates in men with systolic heart failure: an observational study. Journal of Physiotherapy 2013 Sep;59(3):177-187. [doi: 10.1016/s1836-9553(13)70182-6]
21. Andrianopoulos V, Wouters EF, Pinto-Plata VM, Vanfleteren LE, Bakke PS, Franssen FM, et al. Prognostic value of variables derived from the six-minute walk test in patients with COPD: Results from the ECLIPSE study. Respir Med 2015 Sep;109(9):1138-1146 [FREE Full text] [doi: 10.1016/j.rmed.2015.06.013] [Medline: 26143282]
22. Lamberti N, López-Soto PJ, Guerzoni F, Napoli N, Gasbarro V, Zamboni P, et al. Changes in exercise capacity and risk of all-cause mortality in patients with peripheral artery disease: a 10-year retrospective cohort study. Intern Emerg Med 2020 Mar 21;15(2):289-298. [doi: 10.1007/s11739-019-02176-3] [Medline: 31435898]
23. McDermott MM, Kibbe MR, Guralnik JM, Ferrucci L, Criqui MH, Domanchuk K, et al. Durability of benefits from supervised treadmill exercise in people with peripheral artery disease. JAHA 2019 Jan 08;8(1):e009380. [doi: 10.1161/jaha.118.009380]
24. McDermott MM, Tian L, Criqui MH, Ferrucci L, Conte MS, Zhao L, et al. Meaningful change in 6-minute walk in people with peripheral artery disease. J Vasc Surg 2021 Jan;73(1):267-276.e1. [doi: 10.1016/j.jvs.2020.03.052] [Medline: 32335305]
25. Hageman D, Fokkenrood H, Gommans L, van den Houten MM, Teijink J. Supervised exercise therapy versus home-based exercise therapy versus walking advice for intermittent claudication. Cochrane Database Syst Rev 2018 Apr 06;4:CD005263-CD005131 [FREE Full text] [doi: 10.1002/14651858.CD005263.pub4] [Medline: 29627967]
26. McDermott MM. Exercise training for intermittent claudication. J Vasc Surg 2017 Nov;66(5):1612-1620 [FREE Full text] [doi: 10.1016/j.jvs.2017.05.111] [Medline: 28874320]
27. Hotta K, Batchelor WB, Graven J, Dahya V, Noel TE, Ghai A, et al. Daily passive muscle stretching improves flow-mediated dilation of popliteal artery and 6-minute walk test in elderly patients with stable symptomatic peripheral artery disease. Cardiovasc Revasc Med 2019 Aug;20(8):642-648 [FREE Full text] [doi: 10.1016/j.carrev.2019.05.003] [Medline: 31171470]
28. Treat-Jacobson D, Lindquist RA, Witt DR, Kirk LN, Schorr EN, Bronas UG, et al. The PADQOL: Development and validation of a PAD-specific quality of life questionnaire. Vasc Med 2012 Nov 26;17(6):405-415. [doi: 10.1177/1358863x12466708]
29. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the mobile application rating scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
30. Saner H, van der Velde E. eHealth in cardiovascular medicine: A clinical update. Eur J Prev Cardiol 2016 Oct 26;23(2 suppl):5-12. [doi: 10.1177/2047487316670256] [Medline: 27892420]

31. Kao C, Liebovitz DM. Consumer mobile health apps: Current state, barriers, and future directions. PM R 2017 May 18;9(5S):S106-S115. [doi: 10.1016/j.pmrj.2017.02.018] [Medline: 28527495]
32. Miller DP, Weaver KE, Case LD, Babcock D, Lawler D, Denizard-Thompson N, et al. Usability of a novel mobile health iPad app by vulnerable populations. JMIR Mhealth Uhealth 2017 Apr 11;5(4):e43 [FREE Full text] [doi: 10.2196/mhealth.7268] [Medline: 28400354]
33. Geoghegan C, Nido V, Bemden AB, Hallinan Z, Jordan L, Kehoe LS, et al. Learning from patient and site perspectives to develop better digital health trials: Recommendations from the Clinical Trials Transformation Initiative. Contemp Clin Trials Commun 2020 Sep;19:100636 [FREE Full text] [doi: 10.1016/j.conctc.2020.100636] [Medline: 32913915]
34. Karageorgos G, Andreadis I, Psychas K, Mourkousis G, Kiourti A, Lazzi G, et al. The promise of mobile technologies for the health care system in the developing world: a systematic review. IEEE Rev. Biomed. Eng 2019;12:100-122. [doi: 10.1109/rbme.2018.2868896]
35. Guetterman TC, Sakakibara R, Baireddy S, Kron FW, Scerbo MW, Cleary JF, et al. Medical students' experiences and outcomes using a virtual human simulation to improve communication skills: Mixed methods study. J Med Internet Res 2019 Nov 27;21(11):e15459 [FREE Full text] [doi: 10.2196/15459] [Medline: 31774400]
36. Williams JS, Walker RJ, Smalls BL, Hill R, Egede LE. Patient-centered care, glycemic control, diabetes self-care, and quality of life in adults with Type 2 diabetes. Diabetes Technol Ther 2016 Oct;18(10):644-649 [FREE Full text] [doi: 10.1089/dia.2016.0079] [Medline: 27541872]
37. Quinn CC, Staub S, Barr E, Gruber-Baldini A. Mobile support for older adults and their caregivers: Dyad usability study. JMIR Aging 2019 May 23;2(1):e12276 [FREE Full text] [doi: 10.2196/12276] [Medline: 31518271]
38. Haberlin C, O'Dwyer T, Mockler D, Moran J, O'Donnell DM, Broderick J. The use of eHealth to promote physical activity in cancer survivors: a systematic review. Support Care Cancer 2018 Oct 16;26(10):3323-3336. [doi: 10.1007/s00520-018-4305-z] [Medline: 29909476]
39. van Houwelingen CT, Ettema RG, Antonietti MG, Kort HS. Understanding older people's readiness for receiving telehealth: Mixed-method study. J Med Internet Res 2018 Apr 06;20(4):e123 [FREE Full text] [doi: 10.2196/jmir.8407] [Medline: 29625950]
40. Kruse C, Fohn J, Wilson N, Nunez Patlan E, Zipp S, Mileski M. Utilization barriers and medical outcomes commensurate with the use of telehealth among older adults: Systematic review. JMIR Med Inform 2020 Aug 12;8(8):e20359. [doi: 10.2196/20359]
41. An H, Jones GC, Kang S, Welk GJ, Lee J. How valid are wearable physical activity trackers for measuring steps? Eur J Sport Sci 2017 Apr 02;17(3):360-368. [doi: 10.1080/17461391.2016.1255261] [Medline: 27912681]
42. Leong JY, Wong JE. Accuracy of three Android-based pedometer applications in laboratory and free-living settings. J Sports Sci 2017 Jan 07;35(1):14-21. [doi: 10.1080/02640414.2016.1154592] [Medline: 26950687]
43. Awais M, Mellone S, Chiari L. Physical activity classification meets daily life: Review on existing methodologies and open challenges. Physical activity classification meets daily life: Review on existing methodologies and open challenges. Annu Int Conf IEEE Eng Med Biol Soc. 2015; 2015 Presented at: Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 25–29 August 2015; Milan, Italy p. 5050-5053. [doi: 10.1109/embc.2015.7319526]
44. Wanner M, Richard A, Martin B, Faeh D, Rohrmann S. Associations between self-reported and objectively measured physical activity, sedentary behavior and overweight/obesity in NHANES 2003-2006. Int J Obes (Lond) 2017 Jan 28;41(1):186-193. [doi: 10.1038/ijo.2016.168] [Medline: 27677618]
45. Ball TJ, Joy EA, Gren LH, Shaw JM. Concurrent validity of a self-reported physical activity “vital sign” questionnaire with adult primary care patients. Prev. Chronic Dis 2016 Feb 04;13:E16. [doi: 10.5888/pcd13.150228]
46. Yin K, Laranjo L, Tong HL, Lau AY, Kocaballi AB, Martin P, et al. Context-aware systems for chronic disease patients: Scoping review. J Med Internet Res 2019 Jun 17;21(6):e10896 [FREE Full text] [doi: 10.2196/10896] [Medline: 31210138]
47. Golledge J, Leicht AS, Yip L, Rowbotham SE, Pinchbeck J, Jenkins JS, et al. Relationship between disease specific quality of life measures, physical performance, and activity in people with intermittent claudication caused by peripheral artery disease. Eur J Vasc Endovasc Surg 2020 Jun;59(6):957-964 [FREE Full text] [doi: 10.1016/j.ejvs.2020.02.006] [Medline: 32336621]
48. Salisbury DL, Whipple MO, Burt M, Brown RJ, Hirsch A, Foley C, et al. Translation of an evidence-based therapeutic exercise program for patients with peripheral artery disease. J Vasc Nurs 2018 Mar;36(1):23-33 [FREE Full text] [doi: 10.1016/j.jvn.2017.09.003] [Medline: 29452626]
49. Novakovic M, Jug B, Lenasi H. Clinical impact of exercise in patients with peripheral arterial disease. Vascular 2017 Aug 09;25(4):412-422. [doi: 10.1177/1708538116678752] [Medline: 28256934]

Abbreviations

ABI: ankle-brachial index
CAD: coronary artery disease
ESC: European Society of Cardiology

mHealth: mobile health
PAD: peripheral arterial disease
PAD-QoL: peripheral arterial disease–related quality of life CVD: cardiovascular disease
PWD: pain-free walking distance
QoL: quality of life
SET: supervised exercise training

Edited by L Buis; submitted 09.09.20; peer-reviewed by SGS Shah, J Santos; comments to author 09.10.20; revised version received 16.11.20; accepted 24.02.21; published 16.08.21.

Please cite as:
Paldán K, Steinmetz M, Simanovski J, Rammos C, Ullrich G, Jánosi RA, Moebus S, Rassaf T, Lortz J
Supervised Exercise Therapy Using Mobile Health Technology in Patients With Peripheral Arterial Disease: Pilot Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(8):e24214
URL: https://mhealth.jmir.org/2021/8/e24214
doi:10.2196/24214
PMID:34398800

Original Paper

Usefulness of Smartphone Apps for Improving Nutritional Status of Pancreatic Cancer Patients: Randomized Controlled Trial

Jiyoung Keum[1,2], MD; Moon Jae Chung[2], MD, PhD; Youngin Kim[3,4], MD; Hyunyoung Ko[4], MSc; Min Je Sung[2,5], MD; Jung Hyun Jo[2], MD; Jeong Youp Park[2], MD, PhD; Seungmin Bang[2], MD, PhD; Seung Woo Park[2], MD, PhD; Si Young Song[2], MD, PhD; Hee Seung Lee[2], MD, PhD

[1]Division of Gastroenterology, Department of Internal Medicine, Ewha Womans University College of Medicine, Seoul, Republic of Korea

[2]Division of Gastroenterology, Department of Internal Medicine, Institute of Gastroenterology, Yonsei University College of Medicine, Seoul, Republic of Korea

[3]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea

[4]Noom Korea, Inc, Seoul, Republic of Korea

[5]Department of Gastroenterology, CHA Bundang Medical Center, CHA University, Seongnam, Republic of Korea

Corresponding Author:
Hee Seung Lee, MD, PhD
Division of Gastroenterology, Department of Internal Medicine
Institute of Gastroenterology
Yonsei University College of Medicine
50 Yonsei-ro
Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 1050586865
Email: LHS6865@yuhs.ac

Abstract

Background: Approximately 80% of pancreatic ductal adenocarcinoma (PDAC) patients suffer from anorexia, weight loss, and asthenia. Most PDAC patients receive chemotherapy, which often worsens their nutritional status owing to the adverse effects of chemotherapy. Malnutrition of PDAC patients is known to be associated with poor prognosis; therefore, nutritional management during chemotherapy is a key factor influencing the outcome of the treatment. Mobile apps have the potential to provide readily accessible nutritional support for patients with PDAC.

Objective: We aimed to evaluate the efficacy of a mobile app–based program, Noom, in patients receiving chemotherapy for PDAC.

Methods: We prospectively enrolled 40 patients who were newly diagnosed with unresectable PDAC from a single university-affiliated hospital in South Korea, and randomly assigned them into a Noom user group (n=20) and a non-Noom user group (n=20). The 12-week in-app interventions included meal and physical activity logging as well as nutritional education feedback from dietitians. The non-Noom user group did not receive any nutrition intervention. The primary outcomes were the changes in the nutritional status and quality of life (QoL) from the baseline to 12 weeks. The secondary outcomes included the changes in the skeletal muscle index (SMI) from the baseline to 12 weeks. The European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Core Questionnaire (QLQ-C30) and the Patient-Generated Subjective Global Assessment (PG-SGA) were used as paper questionnaires to assess the QoL and nutritional status of the patients. Intention-to-treat and per-protocol analyses were conducted. Regarding the study data collection time points, we assessed the nutritional status and QoL at the baseline (T0), and at 4 (T1), 8 (T2), and 12 (T3) weeks. Abdominal computed tomography (CT) imaging was conducted at the baseline and after 8 weeks for tumor response and SMI evaluation. The skeletal muscle area (cm^2) was calculated using routine CT images. The cross-sectional areas (cm^2) of the L3 skeletal muscles were analyzed.

Results: Between February 2017 and January 2018, 48 patients were assessed for eligibility. Totally 40 patients with pancreatic cancer were included by random allocation. Only 17 participants in the Noom user group and 16 in the non-Noom user group completed all follow-ups. All the study participants showed a significant improvement in the nutritional status according to the PG-SGA score regardless of Noom app usage. Noom users showed statistically significant improvements on the global health

status (GHS) and QoL scales compared to non-Noom users, based on the EORTC QLQ (*P*=.004). The SMI decreased in both groups during chemotherapy (Noom users, 49.08±12.27 cm^2/m^2 to 46.08±10.55 cm^2/m^2; non-Noom users, 50.60±9.05 cm^2/m^2 to 42.97±8.12 cm^2/m^2). The decrement was higher in the non-Noom user group than in the Noom user group, but it was not statistically significant (-13.96% vs. -3.27%; *P*=.11).

Conclusions: This pilot study demonstrates that a mobile app–based approach is beneficial for nutritional and psychological support for PDAC patients receiving chemotherapy.

Trial Registration: ClinicalTrials.gov NCT04109495; https://clinicaltrials.gov/ct2/show/NCT04109495.

(JMIR Mhealth Uhealth 2021;9(8):e21088) doi:10.2196/21088

KEYWORDS

pancreatic ductal adenocarcinoma; mobile app; nutritional support; quality of life; chemotherapy

Introduction

Cancer cachexia is associated with poor therapeutic response, treatment-related adverse events, and low quality of life (QoL) in pancreatic ductal adenocarcinoma (PDAC) patients [1]. Approximately 80% of PDAC patients suffer from a wasting syndrome known as "cancer anorexia-cachexia syndrome," which is characterized by anorexia, weight loss, asthenia, and poor prognosis [2-7]. Patients with PDAC have high risk of nutritional malabsorption and metabolic problems compared to patients with other types of cancers, as the pancreas plays a crucial role in exocrine and endocrine functions [8-11].

In addition, almost 80% of PDAC patients are not operable at diagnosis and receive palliative chemotherapy [12]. Chemotherapy aggravates anorexia, nausea, vomiting, and abdominal pain, which can also affect the patients' QoL and nutritional status. Unfortunately, studies on nutritional evaluation and management of PDAC patients during chemotherapy are insufficient. Malnutrition of PDAC patients is associated with chemotherapy-induced toxicity, low adherence to anticancer treatment, as well as poor QoL and survival [2,4,6,13]. Therefore, research on the nutritional management for PDAC patients during active treatment should be considered.

Digital health care systems, especially mobile apps, have the potential to provide readily accessible nutritional and psychological support for cancer patients [14-23]. To date, there has been no randomized controlled clinical trial to evaluate the effectiveness of app-based programs targeting patients with PDAC undergoing chemotherapy. The main purpose of this pilot study was to evaluate the efficacy of mobile app–based supportive care for PDAC patients in the aspects of nutritional status, skeletal muscle index (SMI) change, and QoL.

Methods

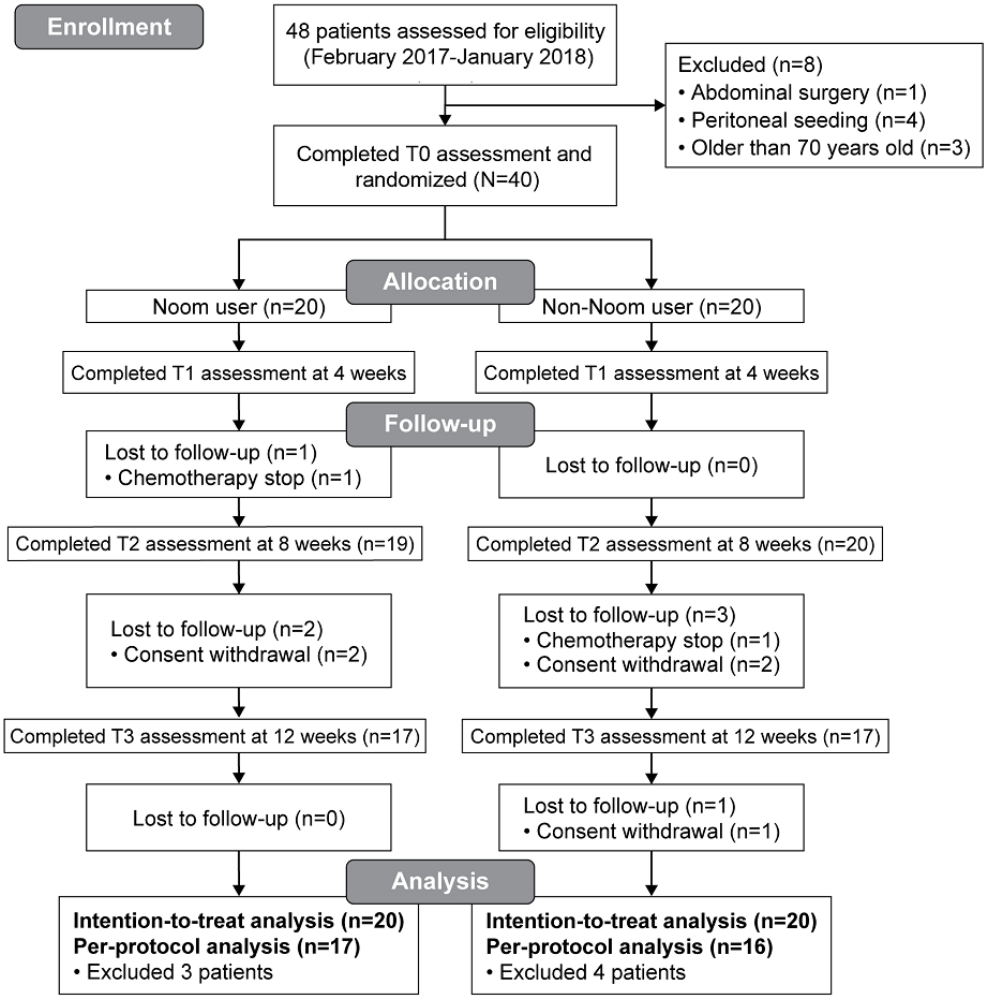
Study Participants

In this randomized controlled trial (RCT) (Trial number NCT 04109495), the study participants were prospectively recruited at a tertiary hospital in South Korea between February 2017 and January 2018. The inclusion criteria were as follows: (1) males or females aged between 20 and 70 years; (2) patients newly diagnosed with PDAC within the last 3 months and slated to receive chemotherapy, (patients) able to access the Internet through their mobile phones; and (4) patients able to read and write Korean. The exclusion criteria were as follows: (1) history of abdominal surgery within the past year and with plans to undergo abdominal surgery; (2) acute illness or infection status (pneumonia, sepsis, shock, etc.); (3) known chronic liver and obstructive pulmonary diseases; (4) known absorption disorder due to gastrointestinal (GI) mucosal disease (ulcerative colitis, Crohn disease, acute and chronic diarrhea, etc.); (5) severe major illness (heart failure, liver failure, kidney failure on hemodialysis, etc.); (6) pregnancy or breastfeeding; (7) use of steroids within the past month before recruitment; (8) being diagnosed with or suspected of having peritoneal seeding or GI obstruction; and (9) history of consuming nutritional supplements.

Study Design

This study was a 12-week prospective, single-center, nonblinded, RCT. The clinicians introduced this study to eligible patients in the clinics, and the researchers met interested patients and confirmed their eligibility. After obtaining written informed consent, all patients were randomly assigned in a 1:1 ratio to the Noom user group or non-Noom user group by the clinical research coordinator (Figure 1). Treatment allocation was performed by the randomized permuted block method using random number tables. As this is a pilot study, we set a target sample size of 40 patients considering the rules of thumb. Browne cites a general flat rule to include at least 30 subjects or more to estimate a parameter [24]. Owing to the nature of the intervention, participant details could not be blinded. This study was approved by the institutional review board of the Severance Hospital (Approval number 1-2016-0061).

Figure 1. Flow chart of the study recruitment process.



Interventions

Noom (Noom Inc.) is a mobile app for weight management that is commercially available on Google Playstore and Apple Appstore. Noom has a unique curriculum and human coaching intervention, which is widely used in health and fitness apps [25,26]. We adopted this program to the nutritional and behavior intervention for PDAC patients undergoing chemotherapy. The major goal of nutritional intervention was to encourage caloric intake to maintain the nutritional status and QoC of PDAC patients undergoing chemotherapy. Although the commercialized version of the Noom app was designed for weight loss and healthy dietary intake, we used functions such as food logging, step count, weight logging, and messaging for tracking balanced caloric intake and muscle gain. To achieve this goal, the Noom app offered the following interventions: (1) interactive interface with coach–participant messaging, (2) daily articles for basic health knowledge, (3) food logging with color coding, and (4) automated feedback-based food choices (see Multimedia Appendix 1). The articles provided to patients mainly included basic health knowledge, information on how to organize a diet using calorie density, and exercise, and lifestyle information. Patient-specific feedback was provided by the coach. Participants were asked to log their weight by self-report providing information on their meals and physical activity in the app more than 4 days per week.

The coach, who is a clinical dietitian, provided nutritional intervention based on the following goals: (1) Guide participants to consume more calories than the recommended intake calculated by the Harris–Benedict equation [27,28] with additional disease-related energy requirements [29]. (2) Provide more than four feedbacks per week on nutritional intake. (3) Check the participants' step counts and exercise logs once a week to promote light physical activity. Noom aims to provide nutritional support for PDAC patients through self-management by monitoring their meals and in-app activities. The clinical research coordinator helped the study participants download the Noom app onto their mobile phones and register themselves on the app. A unique username was generated with a personal password. The study participants did not need to pay for accessing Noom.

The participants in both the groups answered paper questionnaires in the presence of the clinical research coordinator and underwent blood tests at the baseline, and at 4, 8, and 12 weeks. The questionnaire items included the gender, age, body weight, type of diagnosed digestive disease, treatment method, status of oral nutritional supplements, and past medical history of the patients. The European Organization for Research

and Treatment of Cancer (EORTC) Quality of Life Core Questionnaire (QLQ-C30, version 3.0) and the Patient-Generated Subjective Global Assessment (PG-SGA) were also used to assess the QoL and nutritional status of the patients.

App activity was calculated as the summation of the recorded events, including meals, exercise, weight input, number of messages, and step counts. Participants with app activity for more than 9 weeks were defined to be "above average users" (10/40), and participants with app activity for less than 9 weeks were defined to be "below average users" (7/40). A 9-week period was determined based on the median value of the app users' activity.

Comparator

As opposed to the Noom user group, the non-Noom user group did not have access to Noom. This group did not receive any nutrition intervention and attended only study assessments. They received the chemotherapeutic agent as usual. To reduce the bias related to the usually prescribed appetite stimulant, the subjects in both the groups who showed no statistical differences received the same dosage of the appetite stimulant. In this study, patients diagnosed with pancreatic cancer received chemotherapy without regular nutritional intervention.

Variables

The primary objective was to investigate the changes in the QoL or nutritional status, which were calculated from the EORTC QLQ-C30 and the PG-SGA score, over time according to Noom usage. The EORTC QLQ-C30 is a 30-item cancer-specific questionnaire that incorporates 5 functional scales (physical, role, cognitive, emotional, and social), 3 symptom scales (fatigue, pain, and nausea/vomiting), a global health status (GHS) and QoL scale, several single items assessing additional symptoms commonly reported by cancer patients (dyspnea, loss of appetite, insomnia, constipation, and diarrhea), and the financial impact of the disease [30]. A higher score on the GHS and QoL scales indicates a good QoL, but a higher score on the symptom scales indicates poor QoL.

The PG-SGA is a scoring method for nutritional measurement integrating body weight, food intake, nutritional difficulties and activities, and therapeutic information provided by physicians. A score ≥9 indicates a critical need for nutritional intervention [31]. Several studies have verified that the PG-SGA is a reliable and valid assessment of the nutritional status of cancer patients; therefore, we used the PG-SGA scale as a nutritional status assessment tool [14,32-34]. A trained nurse assessed all the PG-SGA scores to maintain consistency in the test results.

The secondary objective was to observe changes in the SMI according to Noom usage. We evaluated whether the SMI was associated with Noom usage at the baseline and during the follow-up period. The skeletal muscle area (cm^2) was calculated using routine computed tomography (CT) images through the picture archiving and communication system (PACS), an image system using Image J software (US National Institutes of Health) [35,36]. Cross-sectional areas (cm^2) of the L3 skeletal muscles were analyzed using Image J. At the L3 level, the field of view included the psoas, paraspinal muscles, and abdominal wall muscles. Currently, the most frequently used landmark in the body composition imaging studies for sarcopenia is the L3 level [37]. We segmented the tissues based on the Hounsfield unit of CT scanning using Image J with assistance from a well-trained and an experienced medical doctor. We had previously published several studies using this method [38,39]. The skeletal muscle area was normalized for height (m^2) and calculated as the SMI (cm^2/m^2) [40].

Age, sex, BMI, Eastern Cooperative Oncology Group performance status (ECOG PS), smoking history, tumor extent, tumor size, chemotherapy regimen, and laboratory characteristics (leukocyte, hemoglobin, platelet, albumin, creatinine, and carbohydrate antigen [CA] 19-9) were also examined.

In this study, we attempted to remove the confounding factors such as steroid ingestion, including appetite stimulants, megestrol, and herbs. Clinicians used to prescribe megestrol to stimulate the appetite of cancer patients. To reduce the confounding factors with respect to appetite stimulants, we prescribed the same dose of megestrol (160 mg/day) for all the enrolled patients, except for those who showed good appetite without stimulants. The prescription was confirmed by the clinical judgment of the attending physician.

Statistical Analysis

Data are expressed as the median, n (%), or n, as appropriate. Variables were compared using Chi-square tests or Fisher exact tests for categorical data and Student *t* tests for continuous variables to evaluate the statistical significance of the differences in the baseline characteristics between Noom and non-Noom users. Variables related to the in-app actions of the above average and below average user groups were compared using the Mann-Whitney test. The primary outcomes of the nutritional status and QoL, as measured by the PG-SGA and EORTC QLQ, were analyzed using intention-to-treat analysis and linear mixed models. Intention-to-treat analysis with the last observation carried forward was applied to account for missing data. The secondary outcome of SMI was assessed in a per-protocol analysis, using the Mann-Whitney test. $P<.05$ was considered to indicate statistical significance. Statistical analyses were performed using the SPSS (version 23.0, IBM Corp.).

Results

Patient Characteristics

Between February 2017 and January 2018, 48 patients were assessed for eligibility. A total of 40 patients were enrolled and randomized into 2 groups (Noom users, n=20; non-Noom users, n=20) (Figure 1). After 7 patients dropped out, 17 Noom users and 16 non-Noom users completed all the follow-ups. Attrition was 18% (7/40 participants), including 2 patients (one in the Noom user group and another in the non-Noom user group) who could not continue chemotherapy owing to severe sepsis or progression of disease, and 5 patients (2 in the Noom user group and 3 in the non-Noom user group) who withdrew their informed consent. The baseline variables in Table 1 did not show a significant difference between participants who were

included in the intention-to-treat population (n=40) and per-protocol population (n=33).

The median age was 61.5 years (range 34-78 years), and 25 of the 40 patients (63%) were male. All the recruited patients had unresectable PDAC at the time of diagnosis. The baseline characteristics did not show statistically significant differences between the 2 groups, except for the baseline BMI and hemoglobin. Most of the patients received palliative folinic acid, fluorouracil, irinotecan, and oxaliplatin (FOLFIRINOX) as a first-line chemotherapy (17/20 [85%] of Noom users and 18/20 [90%] of non-Noom users).

Table 1. Baseline characteristics of the study population.

Variables[a]	Intention-to-treat analysis (N=40)			Per-protocol analysis (N=33)		
	Noom users (n=20)	Non-Noom users (n=20)	*P* value[b]	Noom users (n=17)	Non-Noom users (n=16)	*P* value[b]
Age, median, years (range)	62 (45-70)	61 (34-78)	.25	62 (45-70)	61.5 (34-78)	.31
Sex, male, n (%)	13 (65)	12 (60)	.99	10 (58.8)	10 (62.5)	.83
Height (m)	1.63±0.09	1.63±0.09	.7	1.63 ± 0.09	1.63 ± 0.08	.95
Weight (kg)	58.4±7.9	63.5 ± 11.3	.18	58.2 ± 8.4	62.8 ± 10.9	.22
BMI (kg/m^2)	21.91 ± 1.57	23.5 ± 2.72	.03	21.83 ± 1.69	23.46 ± 2.57	.04
ECOG PS,[c] n (%)						
0-1	17 (85)	18 (90)	.99	14 (82.4)	14 (87.5)	.99
2-3	3 (15)	2 (10)		3 (17.6)	2 (12.5)	
Smoking history						
Never	16 (80)	14 (70)	.47	14 (82.4)	14 (87.5)	.71
Former or current	4 (20)	6 (30)		3 (17.6)	2 (12.5)	
DM,[d] n (%)	5 (25)	6 (30)	.72	5 (29.4)	5 (31.3)	.99
SMI (cm^2/m^2)[e]	49.62± 11.62	48.43±9.91	.73	49.08± 12.27	50.60 ± 9.05	.69
WBC,[f] /μL	6310 (5045-7155)	6,975 (6158-8515)	.13	6310 (4925-7340)	6910 (6158-8045)	.26
Hemoglobin, g/dL	12.3 (10.6-13.6)	13.0 (11.9-14.0)	.04	12.3 (10.6-13.2)	13.0 (12.0-13.8)	.04
Platelet, $10^3/\mu L$	204.5 (176.5-353.3)	222 (183-289)	.95	196 (178-335)	237 (196-289)	.85
Albumin, g/dL	3.9 (3.5-4.1)	4.1 (3.5-4.3)	.87	3.9 (3.5-4.1)	3.9 (3.3-4.1)	.74
Creatinine, mg/dL	0.77 (0.53-0.91)	0.67 (0.56-0.82)	.27	0.77 (0.55-0.91)	0.69 (0.52-0.82)	.23
Initial CA[g] 19-9, U/mL	829.6 (190-2768)	310.1 (40.8-1734.8)	.1	920 (136.6-2716)	310 (27.9-1516.5)	.07
Elevated initial CA 19-9, U/mL	19 (95)	16 (80)	.34	16 (94.1)	12 (75)	.18
Clinical stage, n (%)						
Borderline resectable	6 (30)	6 (30)	.93	4 (23.5)	4 (25)	.61
Locally advanced	5 (25)	6 (30)		4 (23.5)	6 (37.5)	
Metastatic	9 (45)	8 (40)		9 (52.9)	6 (37.5)	
Tumor size, cm	4.2±1.3	4.2±2.1	.8	4.2±1.3	4.5±2.2	.65
Chemotherapy regimen, n (%)						
FOLFIRINOX[h]	17 (85)	18 (90)	.99	14 (82.4)	14 (87.5)	.99
Gem/Nab-paclitaxel[i]	3 (15)	2 (10)		3 (17.6)	2 (12.5)	

[a]Data are presented as n (%) for categorical variables and as median (interquartile range) or mean±SD for continuous variables.

[b]*P* values were calculated by Student *t* tests for continuous data and Chi-square or Fisher exact tests for categorical data.

[c]ECOG PS: Eastern Cooperative Oncology Group performance status.

[d]DM: diabetes mellitus.

[e]The skeletal muscle index was calculated from the muscle cross-sectional area (cm^2)/height $(m)^2$ of the lumbar muscle.

[f]WBC: white blood cells.

[g]CA: carbohydrate antigen.

[h]FOLFIRINOX: folinic acid, fluorouracil, irinotecan, and oxaliplatin.

[i]Gem/Nab-paclitaxel: gemcitabine/nanoparticle albumin–bound paclitaxel.

The baseline characteristics of the Noom participants, above average and below average users, are shown in Table 2. The values in Table 2 were calculated by per-protocol analysis, as patients who dropped out (n=7) did not complete the entire 9-week period, which is the median value of the app users' activity. The 17 Noom users were divided as follows: 10 above average users and 7 below average users. There were no baseline differences in the sex, age, and baseline BMI between the 2 Noom user groups (Table 2, sex, 59% vs. 60%; age, 62 vs. 62.5 years; baseline BMI, 21.8 vs. 21.5).

Table 2. Baseline characteristics of Noom users.

Variables[a]	Noom users (N=17)	Above average users (n=10)	Below average users (n=7)
Baseline			
Sex, male, n (%)	10 (58.8)	6 (60)	4 (57.1)
Age, median, years (range)	62 (45-70)	62.5 (45-70)	62 (58-65)
Height (m)	1.63±0.09	1.64±0.1	1.61 ± 0.08
Weight (kg)	58.2±8.4	58.32±8.61	58.04±8.75
BMI (kg/m^2)	21.83±1.69	21.54±1.46	22.24±2.02

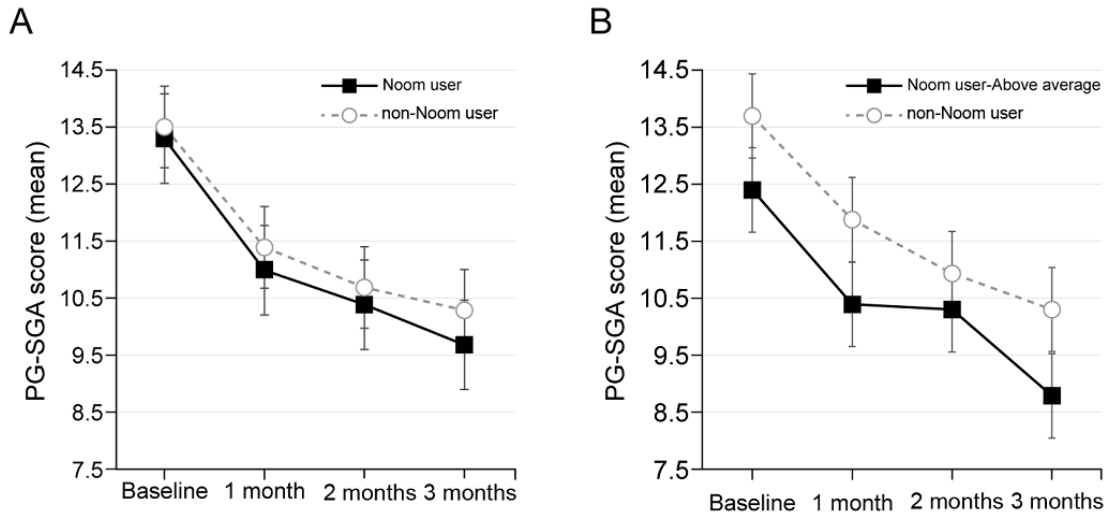
[a]*P* values were calculated by the Mann-Whitney test between the two user groups.

In the intention-to-treat population, 15 of the 20 Noom users (75%) and 14 of 20 non-Noom users (70%) received megestrol, but there was no significant difference in the number of prescribed patients (*P*=.72) and the total dose of the drug (9,440 mg vs. 12,240 mg, *P*=.06). In the per-protocol population, 12 of the 17 Noom users (71%) and 13 of the 16 non-Noom users (81%) received metestrol, but there was no significant difference in the number of prescribed patients (*P*=.69) and the total dose of the drug (11,120 mg vs. 12,320 mg, *P*=.09).

Improvement of Nutritional Status Through Mobile App Usage

In the intention-to-treat analysis, all the study participants showed a significant improvement in the nutritional status according to the PG-SGA score regardless of Noom app usage (Figure 2A, *P*=.001). In the per-protocol analysis, the above average users showed a significant improvement in the PG-SGA score (Figure 2B, *P*=.03).

Figure 2. Change in PG-SGA score according to Noom usage over time. (A) All study participants show an improvement in nutritional status according to the PG-SGA score (*P*=.001). (B) All Noom users show improvement in their nutritional status according to the PG-SGA score regardless of their app activity differences. Above average users showed significant improvement in PG-SGA score (*P*=.03). PG-SGA: Patient-Generated Subjective Global Assessment.



There were significant differences in the total protein and energy intakes between the above average and below average users (Table 3, 1.3 vs. 1g/kg/day, *P*=.02; 25.2 vs. 17.7 kcal/kg/day, *P*=.04). In the per-protocol analysis, 7 of the 10 above average users (70%) met the individual minimum protein intake requirement, and 6 of the 10 above average users (60%) met the individual minimum energy intake requirement. However,

none of the below average users met the minimum recommended daily intake of protein and calories.

The above average users documented meal data more frequently in the Noom app (15.4 meals per week vs. 5.06 meals per week) and showed an increase in their body weight and BMI compared to the below average users (Table 3, body weight, +1.16% vs. -4.43%; BMI, +0.21 vs. -0.81).

Table 3. Effects of Noom app intervention on dietary intake, weight, BMI, and engagement characteristics.

Variables	Noom users (N=17)	Above average users (n=10)	Below average users (n=7)
12 weeks			
Total protein intake(g/kg/day)	1 (0.6-1.4)	1.3 (0.9-1.6)[a]	1 (0.5-1)
Total energy intake(kcal/kg/day)	19.9 (13.9-26.8)	25.2 (17.5-32.7)[b]	17.7 (12.1-20.8)
Weight loss (kg)	-0.66±4.31	0.68±4.80[c]	-2.57±2.76
Weight loss (%)	-1.14±7.55	1.16±8.31	-4.43±5.21
BMI change (kg/m^2)	-0.21±1.43	0.21±1.60[d]	-0.81±0.93
In-app actions[e]			
Meal input frequency (meals per week)	11.15±7.69	15.41±6.76[f]	5.06±3.95
Total exercise input frequency (every 12 weeks)	3.35±8.19	5.6±10.28[g]	0.14±0.38
Articles read (articles/week)	0.98±1.84	1.47±2.3	0.27±0.33
Number of weight inputs (times/week)	0.6±0.72	0.83±0.7	0.27±0.65
Messages to coach (messages/week)	5.14±5.56	6.19±4.35	3.63±7.03
Steps recorded (steps/week)	17,168.23±20,718.02	23,999.58±24,595.55[h]	7,409.17±6,951.79

[a]There was a significant difference in the total protein intake during the 12 weeks between above average and below average users; *P*=.02. The *P* value was calculated by Mann-Whitney tests.

[b]There was a significant difference in the total energy intake during the 12 weeks between above average and below average users; *P*=.04.

[c]There was no significant difference in the changes at 12 weeks between the above average and below average users; *P*=.10.

[d]There was no significant difference in the changes at 12 weeks between the above average and below average users; *P*=.09.

[e]*P* values were calculated by Mann-Whitney tests between the 2 user groups.

[f]All the changes from the baseline to 12 weeks were significant in the above average and below average users; *P*=.007.

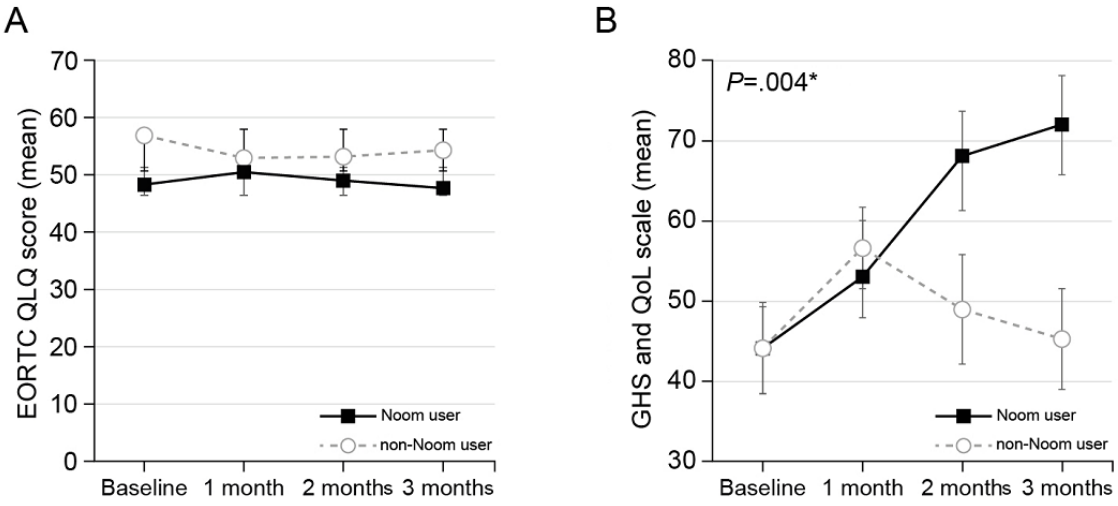
[g]All the changes from the baseline to 12 weeks were significant in the above average and below average users; *P*=.01.

[h]All the changes from the baseline to 12 weeks were significant in the above average and below average users; *P*=.02.

Improvement in QoL Through Mobile App Usage

There was no statistically significant difference in the EORTC QLQ score between the Noom users and non-Noom users (Figure 3A). However, on the GHS and QoL scale, there was a statistically significant improvement in the Noom user group compared to the non-Noom user group during the study period (Figure 3B, *P*=.004).

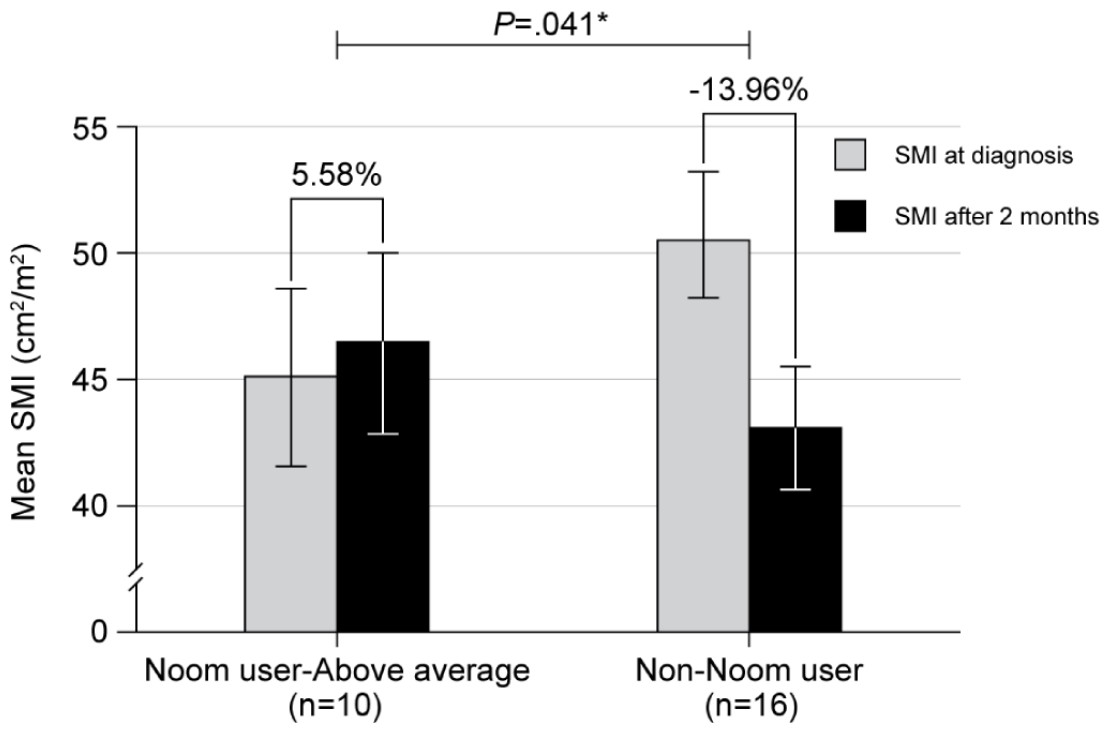
Figure 3. Change in EORTC QLQ-C30 scores according to Noom usage over time. (A) There was no statistically significant difference in EORTC QLQ score between the Noom users and non-Noom users. (B) Noom users showed more statistically significantly improvement on the GHS and QoL scale compared to the non-Noom users over time (*P*=.004). **P* values were calculated by the linear mixed model. EORTC QLQ-C30: European Organization for Research and Treatment of Cancer Quality of Life Core Questionnaire; GHS: global health status; QoL: quality of life.



Skeletal Muscle Change After Mobile App Usage

When receiving chemotherapy, the SMI decreased in both groups (Noom users, 49.08±12.27 cm^2/m^2 to 46.08±10.55 cm^2/m^2; non-Noom users, 50.60±9.05 cm^2/m^2 to 42.97±8.12 cm^2/m^2). The decrement was higher in non-Noom user group than in the Noom user group, but it was not statistically significant (-13.96% vs. -3.27%; *P*=.11). In the per-protocol analysis, there was a statistically significant increment in the SMI of the above average user group compared to the non-Noom user group (Figure 4, +5.58% vs. -13.96%; *P*=.04).

Figure 4. Change in SMI from baseline at 8 weeks. Above average users showed a statistically significant increase in muscle mass after 2 months of Noom usage compared to non-users (*P*=.04). **P* values were calculated by Mann-Whitney tests between the 2 groups. SMI: skeletal muscle index.



Chemotherapy Response According to Mobile App Usage

Furthermore, the overall best response to chemotherapy according to Noom usage was also analyzed in the per-protocol analysis. The overall best response was defined as the best response recorded from the start of chemotherapy until disease progression, recurrence, or the start of new chemotherapy sessions. A decrease in the tumor size was more prominent in the above average users than in the non-Noom users, although this was not statistically significant, possibly owing to the small study population (-15.6% vs. -6.3%; *P*=.17).

There were no differences in terms of the progression-free survival, overall survival, and duration of chemotherapy between the Noom users and non-Noom users.

Discussion

Principal Findings

Mobile apps have been used in the management of chronic diseases such as obesity and hypertension, as well as exercise measurement; they were rarely used directly in hospital care settings. To the best of our knowledge, this is the first study that evaluates the short-term effects of mobile app–based coaching on the change in the nutritional status, SMI, and QoL during PDAC management. The above average Noom users in this pilot study showed statistically significant improvements in their nutritional status. The SMI significantly increased in the above average users compared to the non-Noom users. Moreover, the Noom users showed statistically significant improvements in QoL compared to the non-Noom users based on the GHS and QoL scales of the EORTC QLQ. These findings showed that PDAC patients who receive chemotherapy could be supported by mobile app–based coaching for improving their nutritional and health conditions.

Previous studies have highlighted the importance of supportive care for patients with advanced cancer, including PDAC. Early palliative supportive care led to significant improvements in nutritional statuses, QoL, depression levels, and symptom burden [41-44]. In particular, malnutrition during chemotherapy can induce various adverse effects in humans owing to dysfunction of the intestinal mucosa, decreased immune function, decreased functioning of major organs such as the liver, kidney, and heart, and changes in drug dynamics [10,45-48]. We hypothesized that early nutritional supervision as a supportive treatment could improve the nutritional and psychological conditions of patients with PDAC. In particular, we used a mobile app with human coaching as a nutritional support tool. Even though both groups showed improvements in the PG-SGA score regardless of the frequency of Noom app usage, the above average users showed significant improvements in the PG-SGA scores after using the app. Britton et al reported that the minimum significant difference for the PG-SGA score was 2 points [49]. In this study, the PG-SGA score improved by more than 2 points at 12 weeks in all groups, indicating clinical significance. Therefore, PDAC patients receiving chemotherapy could benefit from using the Noom app, in addition to the usual palliative care.

There have been some attempts at evaluating the efficacy of mobile apps for assisting cancer patients. One randomized study on 114 women with breast cancer who were starting chemotherapy found that e-support program users had better outcomes at 12 weeks for self-efficacy, symptom interference, and QoL compared to users in the control group [15]. Our study also confirmed that QoL improved after using the mobile app. However, the app content used in the two studies differed. Apps used in the Chinese study focused on self-efficacy, social support, and symptom management for patients, whereas the app used in our study focused on nutritional management. Another systematic review also revealed the benefits of app-based programs on the physical activity level, dietary behavior, and health-related QoL in populations diagnosed with solid tumors [50]. Similarly, in our study, the above average users showed better dietary behavior such as inputting their meal data more frequently, with more of them meeting the individual minimum protein and energy intake requirements compared to the below average users. Furthermore, the above average users showed significant improvement in their nutritional status at 12 weeks based on the PG-SGA score. In the present study, the patients had an average PG-SGA score ≥9 points at the baseline, which indicates that almost all participants needed nutritional intervention [2,6,51].

According to previous studies, malnutrition, weight loss, and sarcopenia are risk factors strongly associated with limited tolerance for chemotherapy, short survival times, and poor QoL in PDAC patients. We calculated the SMI using the sarcopenia measurement method. Sarcopenia was measured by CT-based skeletal muscle area assessment, as in many previous studies [52,53]. Loss of skeletal muscle mass is known to be associated with cancer cachexia [54,55]. In the present study, when undergoing chemotherapy, the above average Noom users showed an increment in the SMI, and there was a significant difference between them and the non-Noom users. We assumed that Noom users would be more motivated to improve nutritional intake while monitoring their food intake through the app, and human coaches who are professional nutritionists may have helped Noom users obtain proper nutrition.

A previous retrospective study suggested that early nutritional intervention may affect the overall survival of PDAC patients undergoing chemotherapy [8]. In this study, nutritional intervention included face-to-face dietary consultation with a dietitian. Several studies reported that malnutrition and sarcopenia are factors that are strongly associated with limited tolerance for chemotherapy [3,11]. Despite the increasing evidence demonstrating an association between the nutritional status and clinical outcomes, there is no standard nutritional management tool for PDAC patients undergoing chemotherapy. Although the overall best response to chemotherapy according to Noom usage was not statistically significant, further confirmatory research is needed to achieve better results.

We did not analyze the factors associated with high usage of the app. Although we did not perform statistical analysis, clinical dieticians who coached the patients considered sex as a relevant factor. Female patients or caregivers were considered to use the Noom app more frequently. Further, male patients under 60 years of age were considered to use the Noom app more often.

We believe that further studies on mobile app usage according to the age, sex, and education level of patients and their caregivers, as well as the performance status of the patients will benefit future research on nutritional interventions for cancer patients.

Comparison With Prior Work

Most of the nutritional studies on pancreatic cancer patients undergoing chemotherapy have been performed with face-to-face interventions [56]. Expert dieticians provided essential dietary suggestions and prescribed oral nutritional supplements. These studies have shown positive outcomes, such as improved weight and QoL [8,57,58]. Bauer et al [57] found that cancer patients who received weekly counseling by a dietitian and were advised to consume protein- and energy-dense oral nutritional supplements showed clinically significant improvements in their nutritional status and QoL. Our findings showed a higher improvement in QoL (median change 33.3 vs. 16.7) but lower improvement in nutritional status based on the PG-SGA score (median change 4.5 vs. 9) than those observed in the study by Bauer et al [57], although a direct comparison between the studies was difficult. However, face-to-face studies have limitations in terms of time and space. Mobile health technology is an innovative way to overcome this limitation. Face-to-face interventions can give feedback only on the day of intervention, but with mobile apps, coaches can provide immediate feedback daily based on the patients' meal records. In addition, face-to-face nutritional education is unlikely to be implemented universally owing to time and space constraints. On the other hand, it has the advantage of providing personalized education for patients through their meal records using mobile apps.

Limitations

Our study had several limitations. First, this study could not evaluate the long-term effect of the Noom app, as we conducted this pilot study over a period of only 12 weeks. Further research on the long-term effect of mobile apps on PDAC patients is needed. Second, the study participants were recruited from a single center, and the sample size was small. Furthermore, the requirement of access to the mobile app may have resulted in the participation of a more educated population, potentially limiting the generalization of this study. However, as the number of people familiar with using mobile apps increases over time, it is expected that supportive care using mobile apps could become a promising intervention method. Therefore, additional multicenter-mediated validation is needed to confirm the results of this study. However, one of the strengths of this study was it is the first such study to investigate the use of a mobile app in providing supportive care to PDAC patients. Third, the baseline BMI differed between the Noom users and non-Noom users. The Noom users had a lower BMI at the time of diagnosis compared to that of the non-Noom users. However, the BMI changes in the between the 2 groups did not differ significantly at 12 weeks (P=.99). Therefore, the baseline BMI differences between the 2 groups did not affect the results. Furthermore, SMI reduction was more prominent in the non-Noom users than in the Noom users, despite the higher baseline SMI in the non-Noom users. Fourth, we could not measure the nutritional intake of the non-Noom users owing to the study design; therefore, it was not possible to analyze whether the Noom users consumed more calories and specific nutrients compared to the non-Noom users.

Conclusions

This pilot study demonstrated that a mobile app–based approach for providing nutritional and psychological support could be beneficial for patients with PDAC undergoing chemotherapy. Mobile apps could be useful tools for providing prompt and appropriate nutritional support and monitoring of PDAC patients.

Acknowledgments

The authors wish to thank Noom Inc for supporting this study by providing the mobile app and Boryung Co, Ltd for supplying megestrol acetate.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshots of the Noom app.
[PNG File , 399 KB - mhealth_v9i8e21088_app1.png]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1104 KB - mhealth_v9i8e21088_app2.pdf]

References

1. Fearon K. Cancer cachexia: developing multimodal therapy for a multidimensional problem. Eur J Cancer 2008 May;44(8):1124-1132. [doi: 10.1016/j.ejca.2008.02.033] [Medline: 18375115]

2. Ozola Zalite I, Zykus R, Francisco Gonzalez M, Saygili F, Pukitis A, Gaujoux S, et al. Influence of cachexia and sarcopenia on survival in pancreatic ductal adenocarcinoma: a systematic review. Pancreatology 2015;15(1):19-24. [doi: 10.1016/j.pan.2014.11.006] [Medline: 25524484]
3. Kurita Y, Kobayashi N, Tokuhisa M, Goto A, Kubota K, Endo I, et al. Sarcopenia is a reliable prognostic factor in patients with advanced pancreatic cancer receiving FOLFIRINOX chemotherapy. Pancreatology 2019 Jan;19(1):127-135. [doi: 10.1016/j.pan.2018.11.001] [Medline: 30473464]
4. Basile D, Parnofiello A, Vitale MG, Cortiula F, Gerratana L, Fanotto V, et al. The IMPACT study: early loss of skeletal muscle mass in advanced pancreatic cancer patients. J Cachexia Sarcopenia Muscle 2019 Apr;10(2):368-377 [FREE Full text] [doi: 10.1002/jcsm.12368] [Medline: 30719874]
5. Choi Y, Oh D, Kim T, Lee K, Han S, Im S, et al. Skeletal muscle depletion predicts the prognosis of patients with advanced pancreatic cancer undergoing palliative chemotherapy, independent of body mass index. PLoS One 2015;10(10):e0139749 [FREE Full text] [doi: 10.1371/journal.pone.0139749] [Medline: 26437072]
6. Basile D, Corvaja C, Caccialanza R, Aprile G. Sarcopenia: looking to muscle mass to better manage pancreatic cancer patients. Curr Opin Support Palliat Care 2019 Dec;13(4):279-285. [doi: 10.1097/SPC.0000000000000455] [Medline: 31361630]
7. Gärtner S, Krüger J, Aghdassi AA, Steveling A, Simon P, Lerch MM, et al. Nutrition in pancreatic cancer: a review. Gastrointest Tumors 2016 May;2(4):195-202 [FREE Full text] [doi: 10.1159/000442873] [Medline: 27403414]
8. Trestini I, Carbognin L, Sperduti I, Bonaiuto C, Auriemma A, Melisi D, et al. Prognostic impact of early nutritional support in patients affected by locally advanced and metastatic pancreatic ductal adenocarcinoma undergoing chemotherapy. Eur J Clin Nutr 2018 Mar;72(5):772-779. [doi: 10.1038/s41430-018-0155-5] [Medline: 29581564]
9. Ferrucci LM, Bell D, Thornton J, Black G, McCorkle R, Heimburger DC, et al. Nutritional status of patients with locally advanced pancreatic cancer: a pilot study. Support Care Cancer 2011 Nov;19(11):1729-1734 [FREE Full text] [doi: 10.1007/s00520-010-1011-x] [Medline: 20967470]
10. Gilliland T, Villafane-Ferriol N, Shah K, Shah R, Tran Cao H, Massarweh N, et al. Nutritional and metabolic derangements in pancreatic cancer and pancreatic resection. Nutrients 2017 Mar;9(3):243 [FREE Full text] [doi: 10.3390/nu9030243] [Medline: 28272344]
11. Ryan AM, Prado CM, Sullivan ES, Power DG, Daly LE. Effects of weight loss and sarcopenia on response to chemotherapy, quality of life, and survival. Nutrition 2019;67-68:110539. [doi: 10.1016/j.nut.2019.06.020] [Medline: 31522087]
12. Kamisawa T, Wood LD, Itoi T, Takaori K. Pancreatic cancer. Lancet 2016 Jul;388(10039):73-85. [doi: 10.1016/S0140-6736(16)00141-0] [Medline: 26830752]
13. Tan BH, Birdsell LA, Martin L, Baracos VE, Fearon KC. Sarcopenia in an overweight or obese patient is an adverse prognostic factor in pancreatic cancer. Clin Cancer Res 2009 Nov;15(22):6973-6979 [FREE Full text] [doi: 10.1158/1078-0432.CCR-09-1525] [Medline: 19887488]
14. de Bruin JS, Schuh C, Seeling W, Luger E, Gall M, Hütterer E, et al. Assessing the feasibility of a mobile health-supported clinical decision support system for nutritional triage in oncology outpatients using Arden Syntax. Artif Intell Med 2018 Nov;92:34-42. [doi: 10.1016/j.artmed.2015.10.001] [Medline: 26563776]
15. Zhu J, Ebert L, Liu X, Wei D, Chan SW. Mobile breast cancer e-support program for Chinese Women with breast cancer undergoing chemotherapy (part 2): multicenter randomized controlled trial 21088. JMIR Mhealth Uhealth 2018 Apr;6(4):e104 [FREE Full text] [doi: 10.2196/mhealth.9438] [Medline: 29712622]
16. van der Hout A, van Uden-Kraan CF, Holtmaat K, Jansen F, Lissenberg-Witte BI, Nieuwenhuijzen GAP, et al. Role of eHealth application Oncokompas in supporting self-management of symptoms and health-related quality of life in cancer survivors: a randomised, controlled trial. Lancet Oncol 2020 Jan;21(1):80-94. [doi: 10.1016/S1470-2045(19)30675-8] [Medline: 31838009]
17. Cheng C, Ho RTH, Guo Y, Zhu M, Yang W, Li Y, et al. Development and feasibility of a mobile health-supported comprehensive intervention model (CIMmH) for improving the quality of life of patients with esophageal cancer after esophagectomy: prospective, single-arm, nonrandomized pilot study. J Med Internet Res 2020 Aug;22(8):e18946 [FREE Full text] [doi: 10.2196/18946] [Medline: 32808933]
18. Ahmad AR, Murad HR. The impact of social media on panic during the COVID-19 pandemic in Iraqi Kurdistan: online questionnaire study. J Med Internet Res 2020 May;22(5):e19556 [FREE Full text] [doi: 10.2196/19556] [Medline: 32369026]
19. Young-Afat DA, van GCH, Bruinvels DJ, van DPCC, Witkamp AJ, Sijtsema S, et al. Patients' and health care providers' opinions on a supportive health app during breast cancer treatment: a qualitative evaluation. JMIR Cancer 2016 Jun;2(1):e8 [FREE Full text] [doi: 10.2196/cancer.5334] [Medline: 28410170]
20. Moore TJ, Alsabeeh N, Apovian CM, Murphy MC, Coffman GA, Cullum-Dugan D, et al. Weight, blood pressure, and dietary benefits after 12 months of a Web-based nutrition education program (DASH for health): longitudinal observational study. J Med Internet Res 2008 Dec;10(4):e52 [FREE Full text] [doi: 10.2196/jmir.1114] [Medline: 19073541]
21. Low CA, Dey AK, Ferreira D, Kamarck T, Sun W, Bae S, et al. Estimation of symptom severity during chemotherapy from passively sensed data: exploratory study. J Med Internet Res 2017 Dec;19(12):e420 [FREE Full text] [doi: 10.2196/jmir.9046] [Medline: 29258977]

22. Soh JY, Cha WC, Chang DK, Hwang JH, Kim K, Rha M, et al. Development and validation of a multidisciplinary mobile care system for patients with advanced gastrointestinal cancer: interventional observation study. JMIR Mhealth Uhealth 2018 May;6(5):e115 [FREE Full text] [doi: 10.2196/mhealth.9363] [Medline: 29735478]
23. Kim HJ, Kim SM, Shin H, Jang J, Kim YI, Han DH. A mobile game for patients with breast cancer for chemotherapy self-management and quality-of-life improvement: randomized controlled trial. J Med Internet Res 2018 Oct;20(10):e273 [FREE Full text] [doi: 10.2196/jmir.9559] [Medline: 30578205]
24. Browne RH. On the use of a pilot sample for sample size determination. Stat Med 1995 Sep;14(17):1933-1940. [doi: 10.1002/sim.4780141709] [Medline: 8532986]
25. Pagoto S, Schneider K, Jojic M, DeBiasse M, Mann D. Evidence-based strategies in weight-loss mobile apps. Am J Prev Med 2013 Nov;45(5):576-582. [doi: 10.1016/j.amepre.2013.04.025] [Medline: 24139770]
26. Patel R, Sulzberger L, Li G, Mair J, Morley H, Shing MN, et al. Smartphone apps for weight loss and smoking cessation:quality ranking of 120 apps. N Z Med J 2015 Sep;128(1421):73-76. [Medline: 26370762]
27. Souza MTP, Singer P, Ozorio GA, Rosa VM, Alves MMF, Mendoza López RV, et al. Resting energy expenditure and body composition in patients with head and neck cancer: an observational study leading to a new predictive equation. Nutrition 2018;51-52:60-65. [doi: 10.1016/j.nut.2017.12.006] [Medline: 29605765]
28. Sasaki M, Okamoto H, Johtatsu T, Kurihara M, Iwakawa H, Tanaka T, et al. Resting energy expenditure in patients undergoing pylorus preserving pancreatoduodenectomies for bile duct cancer or pancreatic tumors. J Clin Biochem Nutr 2011 May;48(3):183-186 [FREE Full text] [doi: 10.3164/jcbn.10-83] [Medline: 21562636]
29. Elliott L, Molseed LL, McCallum PD, American Dietetic Association. The Clinical Guide to Oncology Nutrition 2nd Edition. USA: American Dietetic Association; 2006.
30. Fayers P, Bottomley A, EORTC Quality of Life Group, Quality of Life Unit. Quality of life research within the EORTC-the EORTC QLQ-C30. European Organisation for Research and Treatment of Cancer. Eur J Cancer 2002 Mar;38 Suppl 4:S125-S133. [doi: 10.1016/s0959-8049(01)00448-8] [Medline: 11858978]
31. Bauer J, Capra S, Ferguson M. Use of the scored Patient-Generated Subjective Global Assessment (PG-SGA) as a nutrition assessment tool in patients with cancer. Eur J Clin Nutr 2002 Aug;56(8):779-785. [doi: 10.1038/sj.ejcn.1601412] [Medline: 12122555]
32. Solheim TS, Laird BJ, Balstad TR, Stene GB, Bye A, Johns N, et al. A randomized phase II feasibility trial of a multimodal intervention for the management of cachexia in lung and pancreatic cancer. J Cachexia Sarcopenia Muscle 2017 Oct;8(5):778-788 [FREE Full text] [doi: 10.1002/jcsm.12201] [Medline: 28614627]
33. Argilés JM, Anguera A, Stemmler B. A new look at an old drug for the treatment of cancer cachexia: megestrol acetate. Clin Nutr 2013 Jun;32(3):319-324. [doi: 10.1016/j.clnu.2013.01.004] [Medline: 23395103]
34. Lorton CM, Griffin O, Higgins K, Roulston F, Stewart G, Gough N, et al. Late referral of cancer patients with malnutrition to dietitians: a prospective study of clinical practice. Support Care Cancer 2020 May;28(5):2351-2360. [doi: 10.1007/s00520-019-05042-2] [Medline: 31485981]
35. Gomez-Perez SL, Haus JM, Sheean P, Patel B, Mar W, Chaudhry V, et al. Measuring abdominal circumference and skeletal muscle from a single cross-sectional computed tomography image: a step-by-step guide for clinicians using National Institutes of Health ImageJ. JPEN J Parenter Enteral Nutr 2016 Mar;40(3):308-318 [FREE Full text] [doi: 10.1177/0148607115604149] [Medline: 26392166]
36. Wu C, Chang M, Lyadov VK, Liang P, Chen C, Shih TT, et al. Comparing western and eastern criteria for sarcopenia and their association with survival in patients with pancreatic cancer. Clin Nutr 2019 Apr;38(2):862-869. [doi: 10.1016/j.clnu.2018.02.016] [Medline: 29503056]
37. Lee K, Shin Y, Huh J, Sung YS, Lee IS, Yoon KH, et al. Recent issues on body composition imaging for sarcopenia evaluation. Korean J Radiol 2019 Feb;20(2):205-217 [FREE Full text] [doi: 10.3348/kjr.2018.0479] [Medline: 30672160]
38. Keum J, Lee HS, Kang H, Jo JH, Chung MJ, Park JY, et al. Single-center risk factor analysis for FOLFIRINOX associated febrile neutropenia in patients with pancreatic cancer. Cancer Chemother Pharmacol 2020 Apr;85(4):651-659. [doi: 10.1007/s00280-020-04051-x] [Medline: 32185483]
39. Lee HS, Kim SY, Chung MJ, Park JY, Bang S, Park SW, et al. Skeletal muscle mass predicts poor prognosis in patients with advanced pancreatic cancer undergoing second-line FOLFIRINOX chemotherapy. Nutr Cancer 2019;71(7):1100-1107. [doi: 10.1080/01635581.2019.1597906] [Medline: 30955349]
40. Martin L, Birdsell L, Macdonald N, Reiman T, Clandinin MT, McCargar LJ, et al. Cancer cachexia in the age of obesity: skeletal muscle depletion is a powerful prognostic factor, independent of body mass index. J Clin Oncol 2013 Apr;31(12):1539-1547. [doi: 10.1200/JCO.2012.45.2722] [Medline: 23530101]
41. Temel JS, Greer JA, Muzikansky A, Gallagher ER, Admane S, Jackson VA, et al. Early palliative care for patients with metastatic non-small-cell lung cancer. N Engl J Med 2010 Aug;363(8):733-742. [doi: 10.1056/NEJMoa1000678] [Medline: 20818875]
42. Kavalieratos D, Corbelli J, Zhang D, Dionne-Odom JN, Ernecoff NC, Hanmer J, et al. Association between palliative care and patient and caregiver outcomes: a systematic review and meta-analysis. JAMA 2016 Nov;316(20):2104-2114 [FREE Full text] [doi: 10.1001/jama.2016.16840] [Medline: 27893131]

43. Temel JS, Greer JA, El-Jawahri A, Pirl WF, Park ER, Jackson VA, et al. Effects of early integrated palliative care in patients with lung and GI cancer: a randomized clinical trial. J Clin Oncol 2017 Mar;35(8):834-841 [FREE Full text] [doi: 10.1200/JCO.2016.70.5046] [Medline: 28029308]
44. Vanbutsele G, Pardon K, Van Belle S, Surmont V, De Laat M, Colman R, et al. Effect of early and systematic integration of palliative care in patients with advanced cancer: a randomised controlled trial. Lancet Oncol 2018 Mar;19(3):394-404. [doi: 10.1016/S1470-2045(18)30060-3] [Medline: 29402701]
45. Nitenberg G, Raynard B. Nutritional support of the cancer patient: issues and dilemmas. Crit Rev Oncol Hematol 2000 Jun;34(3):137-168. [doi: 10.1016/s1040-8428(00)00048-2] [Medline: 10838261]
46. Sierzega M, Niekowal B, Kulig J, Popiela T. Nutritional status affects the rate of pancreatic fistula after distal pancreatectomy: a multivariate analysis of 132 patients. J Am Coll Surg 2007 Jul;205(1):52-59. [doi: 10.1016/j.jamcollsurg.2007.02.077] [Medline: 17617332]
47. Schwegler I, von Holzen A, Gutzwiller J, Schlumpf R, Mühlebach S, Stanga Z. Nutritional risk is a clinical predictor of postoperative mortality and morbidity in surgery for colorectal cancer. Br J Surg 2010 Jan;97(1):92-97. [doi: 10.1002/bjs.6805] [Medline: 20013933]
48. Baracos VE. Cancer-associated malnutrition. Eur J Clin Nutr 2018 Sep;72(9):1255-1259. [doi: 10.1038/s41430-018-0245-4] [Medline: 30185853]
49. Britton B, McCarter K, Baker A, Wolfenden L, Wratten C, Bauer J, et al. Eating As Treatment (EAT) study protocol: a stepped-wedge, randomised controlled trial of a health behaviour change intervention provided by dietitians to improve nutrition in patients with head and neck cancer undergoing radiotherapy. BMJ Open 2015 Jul;5(7):e008921 [FREE Full text] [doi: 10.1136/bmjopen-2015-008921] [Medline: 26231757]
50. Kiss N, Baguley BJ, Ball K, Daly RM, Fraser SF, Granger CL, et al. Technology-supported self-guided nutrition and physical activity interventions for adults with cancer: systematic review. JMIR Mhealth Uhealth 2019 Feb;7(2):e12281 [FREE Full text] [doi: 10.2196/12281] [Medline: 30747720]
51. Davidson W, Ash S, Capra S, Bauer J, Cancer Cachexia Study Group. Weight stabilisation is associated with improved survival duration and quality of life in unresectable pancreatic cancer. Clin Nutr 2004 Apr;23(2):239-247. [doi: 10.1016/j.clnu.2003.07.001] [Medline: 15030964]
52. Tan BH, Birdsell LA, Martin L, Baracos VE, Fearon KC. Sarcopenia in an overweight or obese patient is an adverse prognostic factor in pancreatic cancer. Clin Cancer Res 2009 Nov 15;15(22):6973-6979 [FREE Full text] [doi: 10.1158/1078-0432.CCR-09-1525] [Medline: 19887488]
53. Dalal S, Hui D, Bidaut L, Lem K, Del Fabbro E, Crane C, et al. Relationships among body mass index, longitudinal body composition alterations, and survival in patients with locally advanced pancreatic cancer receiving chemoradiation: a pilot study. J Pain Symptom Manage 2012 Aug;44(2):181-191 [FREE Full text] [doi: 10.1016/j.jpainsymman.2011.09.010] [Medline: 22695045]
54. Fearon K, Strasser F, Anker SD, Bosaeus I, Bruera E, Fainsinger RL, et al. Definition and classification of cancer cachexia: an international consensus. Lancet Oncol 2011 May;12(5):489-495. [doi: 10.1016/S1470-2045(10)70218-7] [Medline: 21296615]
55. Baracos VE, Martin L, Korc M, Guttridge DC, Fearon KCH. Cancer-associated cachexia. Nat Rev Dis Primers 2018 Jan;4:17105. [doi: 10.1038/nrdp.2017.105] [Medline: 29345251]
56. Hendifar AE, Petzel MQB, Zimmers TA, Denlinger CS, Matrisian LM, Picozzi VJ, Precision Promise Consortium. Pancreas cancer-sssociated weight loss. Oncologist 2019 May;24(5):691-701 [FREE Full text] [doi: 10.1634/theoncologist.2018-0266] [Medline: 30591550]
57. Bauer JD, Capra S. Nutrition intervention improves outcomes in patients with cancer cachexia receiving chemotherapy--a pilot study. Support Care Cancer 2005 Apr;13(4):270-274. [doi: 10.1007/s00520-004-0746-7] [Medline: 15583950]
58. Vashi P, Popiel B, Lammersfeld C, Gupta D. Outcomes of systematic nutritional assessment and medical nutrition therapy in pancreatic cancer. Pancreas 2015 Jul;44(5):750-755. [doi: 10.1097/MPA.0000000000000336] [Medline: 25872172]

Abbreviations

CA: carbohydrate antigen
CT: computed tomography
ECOG PS: Eastern Cooperative Oncology Group performance status
EORTC: European Organization for Research and Treatment of Cancer
FOLFIRINOX: folinic acid, fluorouracil, irinotecan, and oxaliplatin
GHS: global health status
GI: gastrointestinal
PACS: picture archiving and communication system
PDAC: pancreatic ductal adenocarcinoma
PG-SGA: Patient-Generated Subjective Global Assessment
QLQ-C30: Quality of Life Core Questionnaire

QoL: quality of life
RCT: randomized control trial
SMI: skeletal muscle index

Edited by L Buis; submitted 05.06.20; peer-reviewed by N Kiss, A Ramazan; comments to author 18.08.20; revised version received 20.11.20; accepted 17.06.21; published 31.08.21.

Please cite as:
Keum J, Chung MJ, Kim Y, Ko H, Sung MJ, Jo JH, Park JY, Bang S, Park SW, Song SY, Lee HS
Usefulness of Smartphone Apps for Improving Nutritional Status of Pancreatic Cancer Patients: Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(8):e21088
URL: https://mhealth.jmir.org/2021/8/e21088
doi:10.2196/21088
PMID:34463630

Original Paper

Exercise Management Using a Mobile App in Patients With Parkinsonism: Prospective, Open-Label, Single-Arm Pilot Study

Aram Kim[1], MD; Seo Jung Yun[1], MS; Kwan-Sik Sung[1], MS; Yeonju Kim[1], MS; Ju Young Jo[1], MS; Hanseul Cho[2], MSc; Kyudong Park[3], PhD; Byung-Mo Oh[1,4,5], MD, PhD; Han Gil Seo[1], MD, PhD

[1]Department of Rehabilitation Medicine, Seoul National University College of Medicine, Seoul National University Hospital, Seoul, Republic of Korea

[2]College of Medicine, Kyung Hee University, Seoul, Republic of Korea

[3]School of Information Convergence, Kwangwoon University, Seoul, Republic of Korea

[4]Institute on Aging, Seoul National University, Seoul, Republic of Korea

[5]National Traffic Injury Rehabilitation Hospital, Yangpyeong, Republic of Korea

Corresponding Author:
Han Gil Seo, MD, PhD
Department of Rehabilitation Medicine
Seoul National University College of Medicine
Seoul National University Hospital
101, Daehak-ro, Jongno-gu
Seoul, 03080
Republic of Korea
Phone: 82 220721659
Fax: 82 27437473
Email: hgseo80@gmail.com

Abstract

Background: Patients with parkinsonism have higher inactivity levels than the general population, and this results in increased comorbidities. Although exercise has benefits for motor function and quality of life (QOL) in patients with parkinsonism, these patients face many barriers to exercise participation, such as lack of motivation, fatigue, depression, and time constraints. Recently, the use of mobile apps has been highlighted as a remote exercise management strategy for patients with chronic diseases.

Objective: This study aimed to evaluate the effects of home-based exercise management with a customized mobile app on the exercise amount, physical activity, and QOL of patients with parkinsonism.

Methods: This was a prospective, open-label, single-arm pilot study. The therapist installed the app in the smartphones of the participants and educated them on how to use the app. The therapist developed an individualized multimodal exercise program that consisted of stretching, strengthening, aerobic, balance and coordination, and oral-motor and vocal exercises. Participants were encouraged to engage in an 8-week home-based exercise program delivered through a customized app. The alarm notifications of the app provided reminders to exercise regularly at home. The primary outcome was the exercise amount. The secondary outcomes were assessed using the International Physical Activity Questionnaire (IPAQ), Parkinson's Disease Questionnaire-39 (PDQ-39), and Geriatric Depression Scale (GDS). The usability of the customized app was assessed using a self-report questionnaire.

Results: A total of 21 participants with parkinsonism completed the intervention and assessment between September and December 2020 (mean age: 72 years; women: 17/21, 81%; men: 4/21, 19%). The participants reported a significant increase in the total amount of exercise (baseline: mean 343.33, SD 206.70 min/week; 8-week follow-up: mean 693.10, SD 373.45 min/week; $P<.001$) and in the amount of each exercise component, including stretching, strengthening, balance and coordination, and oral-motor and vocal exercise after 8 weeks. Analysis of the secondary outcomes revealed significant improvements in the IPAQ ($P=.006$), PDQ-39 ($P=.02$), and GDS ($P=.04$) scores. The usability of the program with the mobile app was verified based on the positive responses such as "intention to use" and "role expectation for rehabilitation."

Conclusions: Exercise management with a customized mobile app may be beneficial for improving exercise adherence, physical activity levels, depression management, and QOL in patients with parkinsonism. This remotely supervised technology-based, reinforcing, and multimodal exercise management strategy is recommended for use in patients with parkinsonism. In addition, this program proved useful as an alternative exercise management strategy during the COVID-19 pandemic when patients with

Parkinson disease were less physically active than before and showed aggravation of symptoms. However, additional clinical trials are needed to evaluate the efficacy of this exercise program in a large population and to confirm its disease-modifying effects.

(JMIR Mhealth Uhealth 2021;9(8):e27662) doi:10.2196/27662

KEYWORDS

Parkinsonian disorders; exercise; mobile apps; mhealth; Parkinson

Introduction

Parkinson disease (PD) is a progressive neurodegenerative disease characterized by motor symptoms including tremors, rigidity, and bradykinesia, and nonmotor symptoms such as depression and cognitive impairment [1]. The physical activity level of patients with PD is approximately one-third of the general population because of their physical, cognitive, and emotional impairments. A previous study investigated the determinants of physical inactivity in detail and reported that disease severity, walking impairments, and disability in daily life are factors associated with physical inactivity [2]. Physical inactivity results in increased comorbidities such as cardiovascular events, diabetes mellitus, cancer, and osteoporosis [3]. This is also probably true in patients with atypical parkinsonism, such as multiple system atrophy (MSA), progressive supranuclear palsy (PSP), and corticobasal degeneration (CBD), which show clinical symptoms similar to those of PD and whose differential diagnosis can be challenging in the early disease stages [4].

Numerous studies have shown the positive effects of physical activity in terms of preventing depression and cognitive decline, as well as for improving strength, balance, and flexibility in patients with PD [5-7]. With respect to the type of exercise, considerable evidence suggests the benefits of daily walking, strength training, and Tai Chi. In addition, swimming, cycling, and dancing are reported to have physical benefits. Although there is insufficient evidence to prove the superiority of one type of exercise over others, owing to heterogeneity in the details of the exercise and outcome variables in previous studies, regular exercise is known to improve motor function and quality of life (QOL) in patients with PD regardless of the exercise type [8]. Tomlinson et al [9] reviewed the short-term benefits of a physical intervention program on gait, balance, and disability outcomes, and Mak et al [10] reviewed the long-term benefits of exercise on motor symptoms and physical function parameters such as muscle strength, aerobic capacity, gait impairment, balance, and fall risk. In particular, long-term improvements caused by exercise may indicate a disease-modifying effect, which has been proven with respect to progressive resistance, aerobic, and balance training in human and animal model studies [11,12]. Although only a few studies have been published on the effect of exercise on atypical parkinsonism, the benefits of several types of exercise have been reported in patients with MSA, PSP, and CBD [13-15].

Exercise is recommended at all disease stages, regardless of active or sedentary states [7,16]. Even active patients face several barriers (eg, lack of motivation, fatigue, depression, and time constraints) to exercise participation, especially to long-term exercise maintenance [17]. An effective exercise management strategy for patients with PD needs to be focus on providing motivational and enjoyable experiences to facilitate regular exercise. With respect to the type of exercise management strategy, previous studies comparing center-based and home-based exercise programs in older adults have found that center-based exercises are superior in the short term and home-based exercises are superior in the long term with respect to adherence [18]. A previous study reported that minimally supervised home-based programs using collaborative goal setting between therapists and patients, exercise diary records, and intermittent follow-ups can improve adherence and achieve sustained improvements in patients with PD [19]. Recently, the use of specially developed mobile apps has been highlighted as a remote exercise management strategy for patients with chronic diseases. Owing to the increasing life expectancy in patients with chronic diseases, maintaining QOL is important and management using mobile apps has been proven to improve outcomes in this population [20]. In addition, contactless methods of health management are emerging owing to unexpected situations such as the COVID-19 pandemic, and elderly people with chronic disease are the most vulnerable. The use of mobile apps can be an alternative exercise management strategy in patients with PD during the COVID-19 pandemic.

This study aimed to evaluate the effects of a minimally supervised home-based exercise program delivered through a customized mobile app on the exercise amount, physical activity, emotional well-being, and QOL in patients with parkinsonism. We also aimed to assess the usability of the mobile app to provide a basis for further studies on the broad application of mobile apps in future.

Methods

Participants

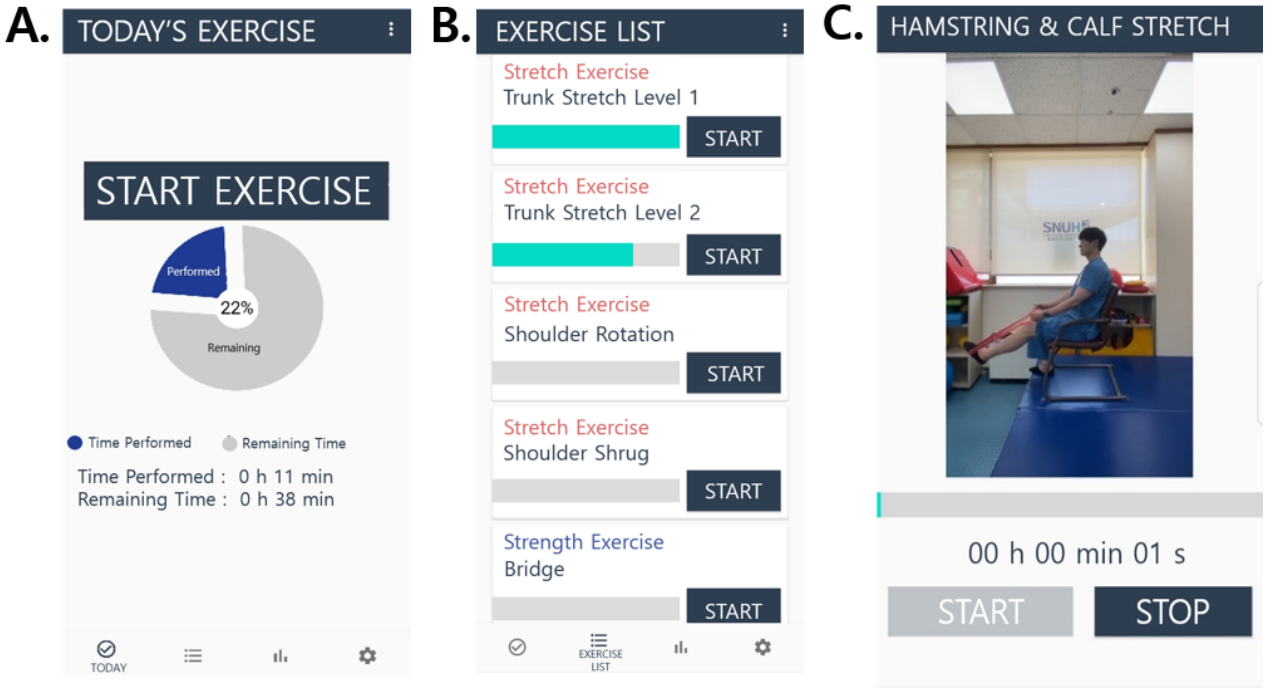
Participants were recruited from the outpatient rehabilitation clinic of a tertiary hospital. The inclusion criteria were as follows: (1) diagnosed with PD or atypical parkinsonism conditions such as MSA, PSP, and CBD; (2) aged above 46 years; and (3) regular participation in a PD exercise program at least once a week. The exclusion criteria were as follows: (1) severe cognitive or physical impairment (Hoehn and Yahr stage 5) interfering with participation in the exercise program during this study and (2) no requirement for exercise management because the amount of recommended exercise was already being performed (>3 h/d in patients at Hoehn and Yahr stages 1-2 and >2 h/d in patients at Hoehn and Yahr stages 2.5-4). The adequate sample size was defined as 24 considering a 20% dropout rate for a pilot study. This study was approved by the institutional

review board of our hospital on August 28, 2020 (approval no. 2007-157-1144), and written informed consent was obtained from all participants. The participants were on common treatment regimens for parkinsonism and maintained their previous exercise programs during the study.

Intervention

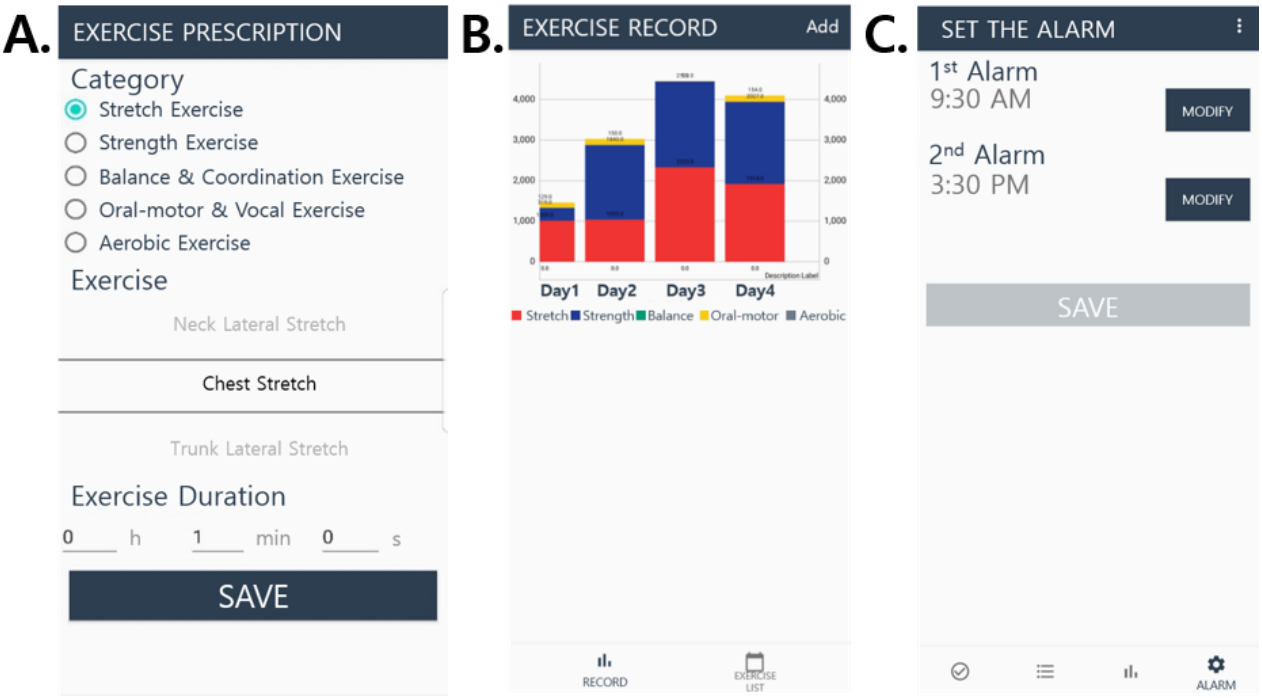
The components of the exercise program were stretching, strengthening, aerobic, balance and coordination, and oral-motor and vocal exercises. The customized app showed the goal (total exercise amount) for the day, a detailed list and the duration for each component, and a video-guided exercise technique recorded by experienced physical and occupational therapists (Figure 1). The alarm notifications for the exercise were initially set by a therapist and later changed by the participants depending on their circumstances. The alarm notifications of the app provided reminders to maintain regular exercise and messages to motivate the patient. In addition, when the patient clicked the button indicating the completion of one exercise component, the app showed the next exercise component to encourage the patient to continue exercising. The app was designed to provide elements of accomplishment and pleasure to promote patient adherence.

Figure 1. Example of an individualized exercise program using a mobile app: (A) goal and degree of completion for the day, (B) list of exercise components, and (C) video-guided exercise technique. (For the convenience of readers, the text in the figure was translated into English [originally in Korean]).



Before starting the exercise program, the therapist installed the app on the personal smartphones of the participants and educated them on how to use the app. The therapist developed an individualized exercise program that consisted of multimodal exercises based on European physiotherapy guidelines for PD [21]. The exercise amount was monitored, and the exercise program and notification time were regularly adjusted by the therapist according to the preference and compliance of each participant (Figure 2).

Figure 2. Example of management by the professional therapist: (A) set of exercise components (type, duration), (B) record of the amount of exercise for monitoring, and (C) set off alarms. (For the convenience of readers, the text in the figure was translated into English [originally in Korean]).



Outcome Measures

Baseline characteristics including the age, sex, diagnosis, disease duration, current medications, Hoehn and Yahr stage, Berg Balance Scale (BBS) score, and timed up-and-go (TUG) test time were assessed. The daily dose of each antiparkinsonian drug was converted into the levodopa equivalent dose (LED), which is a useful indicator of the drug intensity of different medications [22]. The Hoehn and Yahr scale describes the severity of PD from stage 0 through stage 5 [23]. The BBS is a widely used tool for assessing balance performance, in which a higher score indicates better balance ability and a score of 45 points was suggested as the cutoff value for independent ambulation [24]. The TUG test assesses functional mobility, in which a time of <10 seconds means normal mobility and 11–20 seconds is the normal range in elderly patients and patients with disabilities [25].

The outcome measures were assessed at baseline and after using the app for 8 weeks. The assessments included self-completed questionnaires and interviews providing information on exercise amount, physical activity, depression, and QOL.

As the primary outcome, the exercise amount was estimated in terms of the frequency (numbers per week) and duration (minutes per day), and the total amount of exercise was calculated by multiplying the frequency and duration for all exercises and each component. In addition, subjective intensity (Borg scale 6-20) was assessed for the total exercise amount.

To measure physical activity, we used the International Physical Activity Questionnaire (IPAQ), which has been used in the World Health Organization's World Health Survey. The IPAQ questionnaire comprises 7 questions about the frequency and duration of vigorous activity, moderate activity, walking, and sitting. The total physical activity was calculated by multiplying the time (minutes per week) by the intensity (metabolic equivalent of task [MET] unit), and classified as insufficient activity (<600 MET), sufficient activity (600–1499 MET, or vigorous activity for ≥20 minutes on ≥3 days, or moderate activity and walking for ≥30 minutes on ≥2 days), and high activity (1500–2999 MET with ≥3 days of vigorous activity or >2999 MET with ≤2 days of vigorous activity) [26]. The validity of the Korean version of the IPAQ instrument has been proven in the Korean population [27,28].

For evaluating depression, the participants were asked to complete the Geriatric Depression Scale (GDS) Short Form. The GDS is a widely used screening tool for depression in old age. The GDS Short Form contains 15 items. A score of 6 to 7 points was suggested as the cutoff value in Western countries, whereas a score of 10 points was recommended as the cutoff point in a Korean validation study [29].

The QOL of patients with PD was assessed using the Parkinson's Disease Questionnaire-39 (PDQ-39), which is a tool used worldwide and available in many different languages. This tool contains 39 items categorized into 8 dimensions: mobility, activities of daily living, emotional well-being, stigma, social support, cognition, communication, and bodily pain. The total PDQ-39 score is expressed as a percentage out of a total of 100, and a lower score indicates better QOL. The Korean PDQ-39 has been validated in Korean patients with PD [30].

Furthermore, the usability of the customized app was evaluated using a self-report questionnaire including 8 items (symptom improvement, interest, adequate difficulty, physical comfort, stability, satisfaction, intention to use, and role expectations for rehabilitation), with a 7-point Likert-type scale.

Statistical Analysis

Data were analyzed using SPSS software for Windows (version 25.0; IBM Corp). The normality of the data was tested using the Shapiro–Wilk test. To detect a change in the amount of exercise, IPAQ, GDS, and PDQ-39 between baseline (T0) and after 8 weeks (T1), paired *t* tests, and Wilcoxon signed rank tests were used. The level of significance was set at $P<.05$.

Results

Patient Characteristics

Among 28 patients with parkinsonism who were initially screened between September and December 2020, 24 met the inclusion criteria. Two participants withdrew because of fractures associated with a fall event during daily activities, and one participant requested to discontinue the program because of difficulty in using the app. Therefore, a total of 21 participants completed the intervention and assessment.

The characteristics of the participants at baseline are presented in Table 1. The disease severity was moderate to severe, corresponding to Hoehn and Yahr stages 2- 4. In this study, the BBS score and TUG test time ranged from 21 to 56 points and from 13.87 to 7.82 seconds, respectively. All participants were on common dopaminergic treatment (LED: mean 663.94, SD 357.79).

Table 1. Baseline characteristics of participants (N=21).

Participant characteristics	Value
Age (years), mean (SD); range	72.38 (5.77); 62-82
Sex, n (%)	
Female	17 (81)
Male	4 (19)
Diagnosis, n (%)	
Parkinson disease	13 (61.9)
Atypical parkinsonism	8 (38.1)
Disease duration (years), mean (SD); range	8.14 (4.95); 2-18
Hoehn and Yahr scale (stage), n (%)	
2	1 (4.8)
2.5	8 (38.1)
3	8 (38.1)
4	4 (19)
Berg Balance Scale score (points), mean (SD); range	48.38 (9.84); 21-56
Timed up-and-go test time (s), mean (SD); range	13.87 (7.82); 7-42
Levodopa equivalent dose, mean (SD); range[a]	663.94 (357.79); 300-1613

[a]n=18, excluding 3 participants without information on current medications from other medical centers.

Exercise Amount

The participants showed a significant increase of almost 2 times in the total exercise amount per week (T0: mean 343.33, SD 206.70 min/week; T1: mean 693.10, SD 373.45 min/week; $P<.001$) after the study intervention. The amount significantly increased for all exercise components (stretching, strengthening, balance and coordination, oral-motor, and vocal exercises), except for aerobic exercise. A greater increase in the amount of exercise per week was observed in the strengthening component (T1-T0: mean 107.62, SD 83.38 min/week; $P<.001$), followed by the stretching (T1–T0: mean 83.81, SD 188.98 min/week; $P=.04$), balance and coordination (T1-T0: mean 56.90, SD 63.81 min/week; $P<.001$), and oral-motor and vocal (T1-T0: mean 24.38, SD 42.23 min/week; $P=.01$) components. Furthermore, the subjective intensity of exercise (Borg scale 6-20) also increased after 8 weeks of using the app (Table 2).

Table 2. Results for exercise amount, the primary outcome.

Parameter	Baseline (T0)	8 weeks (T1)	Within-individual change (T1-T0)	*P* value
Frequency (number/week), mean (SD)				
Stretching	5.14 (1.93)	5.57 (1.47)	0.43 (2.31)	.53
Strengthening	1.43 (2.54)	5.14 (1.96)	3.71 (2.9)	*<.001*[a]
Aerobic	5.05 (1.71)	5.19 (1.66)	0.14 (1.98)	.75
Balance and coordination	1 (1.73)	3.71 (2.59)	2.71 (3.16)	*.002*
Oral-motor and vocal	2.9 (2.83)	3.9 (2.81)	1 (2.14)	*.045*
Total	5.33 (1.65)[a]	5.86 (1.32)	0.52 (1.66)	.16
Duration (min/day), mean (SD)				
Stretching	17.95 (13.29)	31.52 (29.56)	13.57 (31.71)	.07
Strengthening	3.81 (7.57)	22.62 (14.88)	18.81 (15.64)	*<.001*
Aerobic	42.14 (21.94)	44.29 (23.36)	2.14 (26.3)	.71
Balance and coordination	5.48 (8.65)	18.10 (16.84)	12.62 (17.51)	*.004*
Oral-motor and vocal	7.62 (8.31)	12.00 (13.75)	4.38 (12.15)	.07
Total	61.43 (28.16)	115.48 (54.63)	54.05 (52.86)	*<.001*
Frequency × duration (min/week)				
Stretching	96.86 (81.25)	180.67 (178.40)	83.81 (188.98)	*.04*
Strengthening	15.95 (28.18)	123.57 (89.99)	107.62 (83.38)	*<.001*
Aerobic	211.67 (124.03)	245.24 (169.90)	33.57 (155.07)	.33
Balance and coordination	13.81 (22.47)	70.71 (67.80)	56.90 (63.81)	*.001*
Oral-motor and vocal	34.76 (39.73)	59.12 (59.75)	24.38 (42.23)	*.01*
Total	343.33 (206.70)	693.10 (373.45)	349.76 (344.54)	*<.001*
Intensity (Borg 6–20)	11.86 (1.74)	13.14 (1.42)	1.29 (2.08)	*.02*

[a]Italicized values indicate statistical significance.

Physical Activity, Emotional Well-Being, and QOL

Most participants showed IPAQ scores corresponding to a sufficient activity level at baseline, which indicated that they were active rather than sedentary. We observed a significant increase of almost 2 times in the IPAQ score (T0: mean 1104.17, SD 911.63 MET/week; T1: mean 2027.17, SD 1636.38 MET/week; *P*=.006), classified as high activity, after the intervention. A statistical trend toward a decrease in the time associated with the sedentary state was also observed. In addition, our findings showed that the GDS and PDQ-39 scores were lower at T1 than at T0, indicating significant improvements in depression and QOL (Table 3). Figure 3 shows an overview of the changes in the total amount of exercise, IPAQ score, PDQ-39 score, and GDS score.

Table 3. Results for the secondary outcomes determined using the International Physical Activity Questionnaire, Parkinson Disease Questionnaire-39, and Geriatric Depression Scale.

Parameter	Baseline (T0)	8 weeks (T1)	Within-individual change (T1-T0)	*P* value
IPAQ score[a] (MET/week), mean (SD)				
Vigorous activity	91.43 (418.98)	480.00 (1204.79)	388.57 (1036.39)	.10
Moderate activity	435.24 (578.89)	636.52 (616.20)	201.29 (596.41)	.14
Walking	577.5 (386.68)	910.64 (605.66)	333.14 (488.82)	*.005*
Sedentary	370.48 (196.02)	290 (104.5)	–80.48 (190.17)	.07
Total	*1104.17 (911.63)* [b]	*2027.17 (1636.38)*	*923.00 (1406.38)*	*.006*
PDQ-39[c] score (points), mean (SD)				
Mobility	54.76 (22.76)	49.52 (21.28)	–5.24 (21.88)	.29
Activities of daily living	44.25 (28)	38.69 (27.8)	–5.56 (19.82)	.21
Emotional well-being	47.02 (19.51)	46.63 (22.85)	–0.4 (20.07)	.93
Stigma	37.5 (16.06)	30.65 (14.38)	–6.85 (21.28)	.16
Social support	35.71 (20.44)	37.3 (20.69)	1.59 (21.18)	.74
Cognition	41.67 (19.4)	36.9 (17.67)	–4.76 (10.81)	.06
Communication	37.7 (29.42)	31.75 (27.6)	–5.95 (16.06)	.11
Bodily pain	53.97 (18.93)	36.9 (26.43)	–17.06 (21.49)	*<.001*
Total	44.07 (14.57)	38.54 (14.23)	–5.53 (10.26)	*.02*
GDS[d] score (points)	9.48 (3.42)	7.86 (3.7)	–1.62 (3.25)	.04

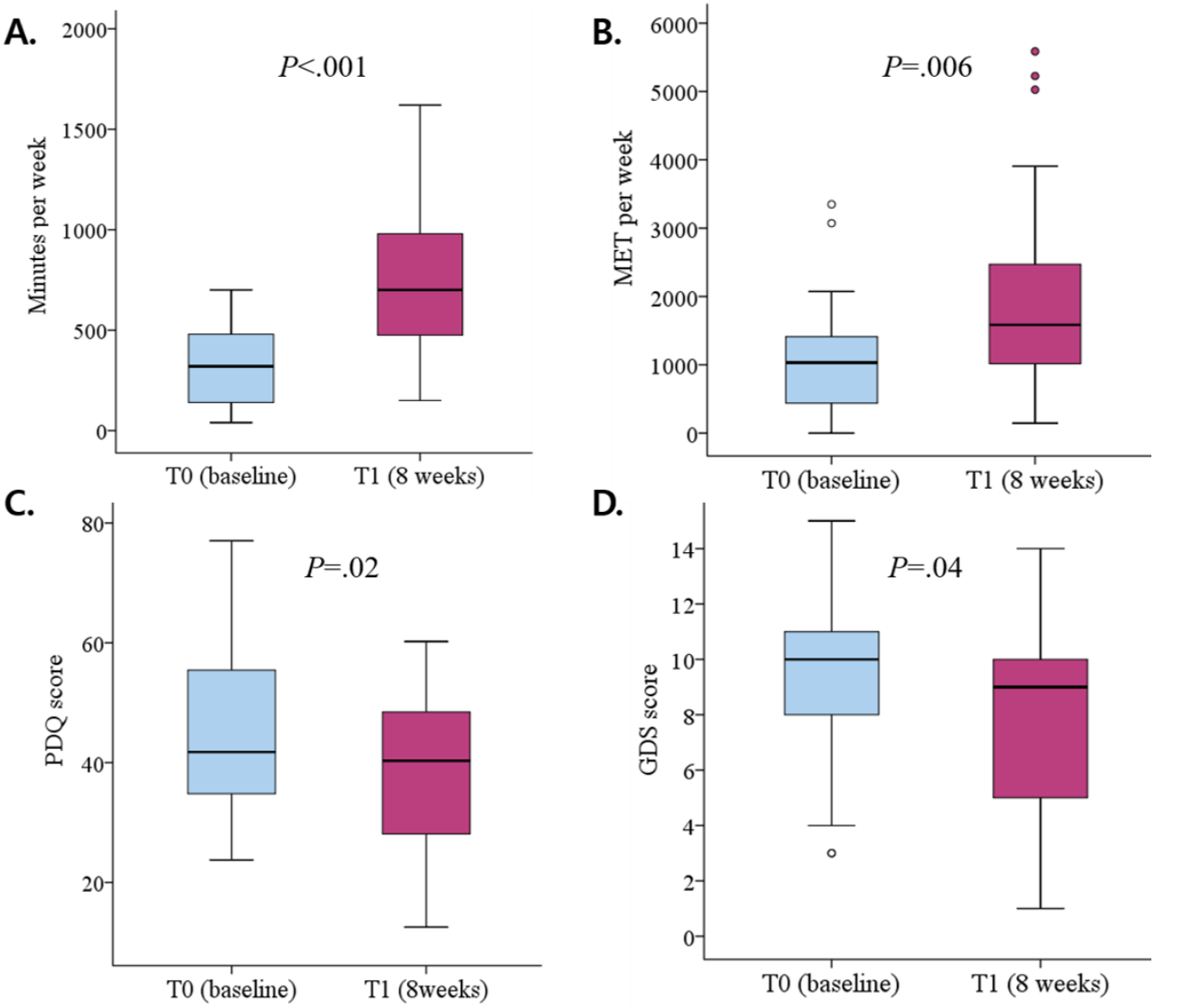
[a]IPAQ: International Physical Activity Questionnaire.

[b]Italicized values indicate statistical significance.

[c]PDQ-39: Parkinson Disease Questionnaire-39.

[d]GDS: Geriatric Depression Scale.

Figure 3. Overview of the changes after participating in the home-based exercise program using a mobile app: (A) total exercise amount, (B) MET score measured using the IPAQ, (C) PDQ-39, and (D) GDS score. GDS: Geriatric Depression Scale; IPAQ: International Physical Activity Questionnaire; MET: metabolic equivalent of task; PDQ: Parkinson Disease Questionnaire.



Usability

With respect to usability according to a 7-point Likert-type scale, the highest score was observed for “intention to use” (mean 6.14, SD 0.77), followed by “role expectation for rehabilitation” (mean 6.10, SD 0.83), “stability” (mean 5.48, SD 1.22), “satisfaction” (mean 5.28, SD 1.03), “interest” (mean 5.24, SD 1.06), “adequate difficulty” (mean 5.24, SD 1.02), “physical comfort” (mean 4.48, SD 1.26), and “symptom improvement” (mean 4.23, SD 0.97).

Discussion

Principal Results

The results of this study indicated a significant increase in the amount of exercise and each component (except for aerobic exercise) after 8 weeks of exercise management using a customized mobile app in patients with parkinsonism. The motivational app had an additional benefit even for participants who had been active before participating in the exercise program. Moreover, we observed significant improvements in physical activity, depression, and QOL at the 8-week follow-up assessment.

The primary outcome of this study was the exercise amount, which represented adherence by patients with parkinsonism to the home-based exercise program with a customized app. Earlier studies have evaluated the effectiveness and usability of tablet-based apps in patients with PD. van der Kolk et al [31] investigated a home-based and remotely supervised aerobic exercise program delivered through a customized tablet-based app (for exercise instructions and monitoring), and they found improvements in the disease severity based on the Unified Parkinson’s Disease Rating Scale (UPDRS) motor score [31]. Siegert et al [32] reported their study protocol for a home-based exercise program using an app installed on a tablet to reinforce health behaviors, established by completing a 3-week center-based exercise program. Furthermore, a few recent studies have investigated the efficacy, feasibility, and safety of using mobile app–based exercise programs. Landers and Ellis [33] reported a single-cohort pilot study on the use of a commercially

available app with a video-guided exercise program for patients with PD. The exercise program was set through an automatic algorithm depending on the function level of the patient and assessed using demographic questions and remote performance-based tests without professional support. The feasibility, safety, and efficacy were proven; however, a high dropout rate was reported. This suggests that "supervised" management by professionals may be necessary for developing an individualized exercise program and for maintaining the patients' adherence to exercise. Ellis et al [34] reported a 12-month randomized controlled pilot study evaluating the effectiveness of mobile health–supported exercise programs compared with only conventional exercise programs in patients with PD. The participants were provided an iPad with an app containing a video-guided exercise prescription, and the exercise prescription was adjusted depending on the remotely monitored compliance. This study found no significant difference between groups in terms of improvements in physical activity, although the mobile health–supported exercise program had more benefits for less active participants (completing <7500 steps/day). Our study can contribute to this growing area of research by exploring the feasibility and efficacy of remotely supervised technology-based reinforcing exercise programs in relatively active patients with parkinsonism. Our findings, although preliminary, suggest that a home-based exercise program using a customized mobile app may be an appropriate strategy for improving compliance in patients with parkinsonism.

In this study, a statistical trend toward an increase in the amount of the aerobic exercise component was observed; however, it was not statistically significant compared to the changes in the other exercise components. A possible explanation may be that aerobic exercise generally does not require a specific technique or instruction. Moreover, as shown in Table 2, aerobic exercise was performed for a relatively longer duration than others, and the participants were already performing aerobic exercise for over 40 min/d at baseline assessment. A long-term and larger study probably would provide significant results on the aerobic exercise component. Conversely, we found the greatest change in the amount of the strengthening component, followed by the stretching and balance components in this study. These results suggested that the customized app was useful for facilitating exercise components that require a specific technique. Many previous studies have recommended multimodal physical activity (a combination of exercise modalities) rather than a single exercise component [8,10]. In particular, strengthening involving the extensor muscles of the hip and trunk, stretching exercises for the flexor and axial muscles, and balance training for individuals with a high fall risk have been recommended [7]. The video-guided instructions set in the app may provide the proper techniques for all the exercise components, resulting in multimodal physical activity.

Another important finding was the improvement in physical activity observed after 2 months of the exercise program. Moreover, a tendency toward decreased sedentary times was also observed. Previous studies have reported several major barriers that lead to sedentary behavior in patients with PD, including lack of motivation, fatigue, depression, low outcome expectation from exercise, lack of time, fear of falling, and low self-efficacy [17,35]. These findings may suggest the need for a management strategy that aims to improve motivation and adherence. Herein, we propose an exercise program with several benefits, as it is cost-effective, provides alarm notifications, is supervised by professionals, offers collaborative goal setting, and allows intermittent monitoring. This strategy focused on aspects such as motivation, accessibility, and compliance, and alarm notifications emerged as the most significant factor in improving adherence. Furthermore, improvements in depression and QOL were observed in this study. The increased physical activity level might have caused a decrease in depressive symptoms and an improvement in QOL of the study participants [5].

Recently, the COVID-19 pandemic has become a major barrier for patients with chronic diseases to obtain medical support. In a recent study, the impact of the COVID-19 pandemic on patients with PD was surveyed [36]. The study found that patients with PD experienced higher COVID-related psychological distress, were less physically active than before, and showed aggravation of symptoms. Although approximately half of the patients were less active than before, no relationship between physical inactivity and psychological distress was established. This type of remotely supervised home-based exercise program using a mobile app may be recommended as an alternative exercise management strategy for patients with parkinsonism during the COVID-19 pandemic.

Limitations

The major limitation of this study was the instability of the customized app. Some errors occurred during login, video streaming, and calculation of the exercise time, which may have interfered with the participation and maintenance of exercise by the patients. Therefore, the cumulative exercise time recorded in the app was used only for monitoring, and the amount of exercise was assessed using self-completed questionnaires. Further work is required to develop an improved app with stable server management and additional attractive features, such as voice instructions, a bigger screen for the elderly, and sharing of exercise data with family and other app users. Another limitation of the study was that we only evaluated the short-term effects of a home-based exercise program with the motivational app. Recent studies reported that long-term improvements caused by exercise may indicate a disease-modifying effect in humans [10,12]. For example, progressive resistance training in patients with PD has been proven to improve motor signs in off-medication UPDRS motor scores. Other studies suggested that an increase in blood oxygen level–dependent signals in basal ganglia circuits and corticomotor excitability results in experience-dependent neuroplasticity [37]. Further studies are warranted to investigate the long-term effect of an exercise program using a customized app in patients with parkinsonism to prove the disease-modifying effects. Lastly, our study included a relatively small sample size, with mixed disease entities in the participants; moreover, there was no control group subjected to a conventional exercise program. In particular, a randomized controlled trial is essential to clarify the effectiveness of the app. Our ongoing follow-up study needs to include a larger sample, more specific disease entity, and control group based on this pilot study.

Conclusions

A home-based exercise program with a customized mobile app has beneficial impacts on adherence to exercise as well as physical activity, depression, and QOL in patients with parkinsonism. We recommend this program as an additional management strategy (characterized as a remotely supervised technology-based, reinforcing, multimodal strategy) for patients with parkinsonism. Moreover, this program can be an alternative exercise management strategy for patients with parkinsonism who are less physically active and are experiencing difficulties in accessing medical services during the COVID-19 pandemic. Additional clinical trials are needed to evaluate the efficacy of this program in a large population and confirm the disease-modifying effects of this exercise program.

Acknowledgments

This study was supported by the Seoul National University Hospital Research Fund (grant 03-2020-2020).

Conflicts of Interest

None declared.

References

1. Armstrong MJ, Okun MS. Diagnosis and treatment of Parkinson disease. JAMA 2020 Feb;323(6):548-560. [doi: 10.1001/jama.2019.22360]
2. van Nimwegen M, Speelman AD, Hofman-van Rossum EJM, Overeem S, Deeg DJH, Borm GF, et al. Physical inactivity in Parkinson's disease. J Neurol 2011 Dec;258(12):2214-2221 [FREE Full text] [doi: 10.1007/s00415-011-6097-7] [Medline: 21614433]
3. Warburton DE, Nicol CW, Bredin SSD. Health benefits of physical activity: the evidence. CMAJ 2006 Mar;174(6):801-809 [FREE Full text] [doi: 10.1503/cmaj.051351] [Medline: 16534088]
4. Poewe W, Wenning G. The differential diagnosis of Parkinson's disease. Eur J Neurol 2002 Nov;9(Suppl 3):23-30. [doi: 10.1046/j.1468-1331.9.s3.3.x] [Medline: 12464118]
5. Wu P, Lee M, Huang T. Effectiveness of physical activity on patients with depression and Parkinson's disease: a systematic review. PLoS One 2017 Jul;12(7):e0181515 [FREE Full text] [doi: 10.1371/journal.pone.0181515] [Medline: 28749970]
6. Díez-Cirarda M, Ojeda N, Peña J, Cabrera-Zubizarreta A, Lucas-Jiménez O, Gómez-Esteban JC, Gómez-Beldarrain, et al. Long-term effects of cognitive rehabilitation on brain, functional outcome and cognition in Parkinson's disease. Eur J Neurol 2017 Oct;25(1):5-12. [doi: 10.1111/ene.13472]
7. van der Kolk NM, King LA. Effects of exercise on mobility in people with Parkinson's disease. Mov Disord 2013 Oct;28(11):1587-1596. [doi: 10.1002/mds.25658]
8. Tomlinson C, Herd C, Clarke C, Meek C, Patel S, Stowe R, et al. Physiotherapy for Parkinson's disease: a comparison of techniques. Cochrane Database Syst Rev 2014 Jun 17(6):CD002815 [FREE Full text] [doi: 10.1002/14651858.CD002815.pub2] [Medline: 24936965]
9. Tomlinson C, Patel S, Meek C, Herd C, Clarke C, Stowe R, et al. Physiotherapy versus placebo or no intervention in Parkinson's disease. Cochrane Database Syst Rev 2013 Sep 10(9):CD002817 [FREE Full text] [doi: 10.1002/14651858.CD002817.pub4] [Medline: 24018704]
10. Mak MK, Wong-Yu IS, Shen X, Chung CL. Long-term effects of exercise and physical therapy in people with Parkinson disease. Nat Rev Neurol 2017 Oct;13(11):689-703. [doi: 10.1038/nrneurol.2017.128]
11. da Silva PGC, Domingues DD, de Carvalho LA, Allodi S, Correa CL. Neurotrophic factors in Parkinson's disease are regulated by exercise: evidence-based practice. J Neurol Sci 2016 Apr;363:5-15. [doi: 10.1016/j.jns.2016.02.017] [Medline: 27000212]
12. Hirsch MA, van Wegen EEH, Newman MA, Heyn PC. Exercise-induced increase in brain-derived neurotrophic factor in human Parkinson's disease: a systematic review and meta-analysis. Transl Neurodegener 2018 Mar;7(1):1-12. [doi: 10.1186/s40035-018-0112-1]
13. Suteerawattananon M, MacNeill B, Protas EJ. Supported treadmill training for gait and balance in a patient with progressive supranuclear palsy. Phys Ther 2002 May;82(5):485-495. [Medline: 11991801]
14. Steffen TM, Boeve BF, Petersen CM, Dvorak L, Kantarci K. Long-term exercise training for an individual with mixed corticobasal degeneration and progressive supranuclear palsy features: 10-year case report follow-up. Phys Ther 2014 Feb;94(2):289-296 [FREE Full text] [doi: 10.2522/ptj.20130052] [Medline: 24114439]
15. Raccagni C, Goebel G, Gaßner H, Granata R, Ndayisaba J, Seebacher B, et al. Physiotherapy improves motor function in patients with the Parkinson variant of multiple system atrophy: a prospective trial. Parkinsonism Relat Disord 2019 Oct;67:60-65. [doi: 10.1016/j.parkreldis.2019.09.026] [Medline: 31621609]
16. van Nimwegen M, Speelman AD, Overeem S, van de Warrenburg BP, Smulders K, Dontje ML, ParkFit Study Group. Promotion of physical activity and fitness in sedentary patients with Parkinson's disease: randomised controlled trial. BMJ 2013 Mar;346(mar01 2):f576-f576 [FREE Full text] [doi: 10.1136/bmj.f576] [Medline: 23457213]

17. Afshari M, Yang A, Bega D. Motivators and barriers to exercise in Parkinson's disease. JPD 2017 Nov;7(4):703-711. [doi: 10.3233/jpd-171173]
18. Ashworth N, Chad K, Harrison E, Reeder B, Marshall S. Home versus center based physical activity programs in older adults. Cochrane Database Syst Rev 2005 Jan 25(1):CD004017 [FREE Full text] [doi: 10.1002/14651858.CD004017.pub2] [Medline: 15674925]
19. Allen NE, Sherrington C, Suriyarachchi GD, Paul SS, Song J, Canning CG. Exercise and motor training in people with Parkinson's disease: a systematic review of participant characteristics, intervention delivery, retention rates, adherence, and adverse events in clinical trials. Parkinsons Dis 2012;2012:854328 [FREE Full text] [doi: 10.1155/2012/854328] [Medline: 22191077]
20. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
21. Keus S, Munneke M, Graziano M, Paltamaa J, Pelosin E, Domingos J. European Physiotherapy guideline for Parkinson's disease. The Netherlands: KNGF/ParkinsonNet. 2014. URL: http://fizioterapeitiem.lv/attachments/article/307/4_eu_guideline_parkinson_201412-development.pdf [accessed 2021-08-02]
22. Tomlinson CL, Stowe R, Patel S, Rick C, Gray R, Clarke CE. Systematic review of levodopa dose equivalency reporting in Parkinson's disease. Mov Disord 2010 Nov 10;25(15):2649-2653. [doi: 10.1002/mds.23429]
23. Hoehn M, Yahr M. Parkinsonism: onset, progression, and mortality. 1967. Neurology 1998 Feb;50(2):318-334. [doi: 10.1212/wnl.50.2.318] [Medline: 9484345]
24. Berg K, Wood-Dauphinee S, Williams J, Maki B. Measuring balance in the elderly: validation of an instrument. Can J Public Health 1992;83(Suppl 2):S7-S11. [Medline: 1468055]
25. Podsiadlo D, Richardson S. J Am Geriatr Soc 1991 Feb;39(2):142-148. [doi: 10.1111/j.1532-5415.1991.tb01616.x] [Medline: 1991946]
26. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International Physical Activity Questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003;35(8):1381-1395. [doi: 10.1249/01.mss.0000078924.61453.fb]
27. Oh JY, Yang YJ, Kim BS, Kang JH. Validity and reliability of Korean version of International Physical Activity Questionnaire (IPAQ) short form. J Korean Acad Fam Med 2007;28(7):532-541.
28. Chun MY. Validity and reliability of Korean version of International Physical Activity Questionnaire short form in the elderly. Korean J Fam Med 2012 May;33(3):144-151 [FREE Full text] [doi: 10.4082/kjfm.2012.33.3.144] [Medline: 22787536]
29. Bae JN, Cho MJ. Development of the Korean version of the Geriatric Depression Scale and its short form among elderly psychiatric patients. J Psychosom Res 2004 Sep;57(3):297-305. [doi: 10.1016/j.jpsychores.2004.01.004] [Medline: 15507257]
30. Park HJ, Sohng KY, Kim S. Validation of the Korean version of the 39-Item Parkinson's Disease Questionnaire (PDQ-39). Asian Nurs Res (Korean Soc Nurs Sci) 2014 Mar;8(1):67-74 [FREE Full text] [doi: 10.1016/j.anr.2014.02.004] [Medline: 25030495]
31. van der Kolk NM, de Vries NM, Kessels RPC, Joosten H, Zwinderman AH, Post B, et al. Effectiveness of home-based and remotely supervised aerobic exercise in Parkinson's disease: a double-blind, randomised controlled trial. The Lancet Neurology 2019 Nov;18(11):998-1008. [doi: 10.1016/s1474-4422(19)30285-6]
32. Siegert C, Hauptmann B, Jochems N, Schrader A, Deck R. ParkProTrain: an individualized, tablet-based physiotherapy training programme aimed at improving quality of life and participation restrictions in PD patients – a study protocol for a quasi-randomized, longitudinal and sequential multi-method study. BMC Neurol 2019 Jun;19(1):1-9. [doi: 10.1186/s12883-019-1355-x]
33. Landers MR, Ellis TD. A mobile app specifically designed to facilitate exercise in Parkinson disease: single-cohort pilot study on feasibility, safety, and signal of efficacy. JMIR Mhealth Uhealth 2020 Oct;8(10):e18985 [FREE Full text] [doi: 10.2196/18985] [Medline: 33016887]
34. Ellis TD, Cavanaugh JT, DeAngelis T, Hendron K, Thomas CA, Saint-Hilaire M, et al. Comparative effectiveness of mHealth-supported exercise compared with exercise alone for people with Parkinson disease: randomized controlled pilot study. Phys Ther 2019 Feb;99(2):203-216. [doi: 10.1093/ptj/pzy131] [Medline: 30715489]
35. Ellis T, Boudreau JK, DeAngelis TR, Brown LE, Cavanaugh JT, Earhart GM, et al. Barriers to exercise in people with Parkinson disease. Phys Ther 2013 May;93(5):628-636 [FREE Full text] [doi: 10.2522/ptj.20120279] [Medline: 23288910]
36. van der Heide A, Meinders MJ, Bloem BR, Helmich RC. The impact of the COVID-19 pandemic on psychological distress, physical activity, and symptom severity in Parkinson's disease. JPD 2020 Oct;10(4):1355-1364. [doi: 10.3233/jpd-202251]
37. Corcos DM, Robichaud JA, David FJ, Leurgans SE, Vaillancourt DE, Poon C, et al. A two-year randomized controlled trial of progressive resistance exercise for Parkinson's disease. Mov Disord 2013 Mar;28(9):1230-1240. [doi: 10.1002/mds.25380]

Abbreviations

BBS: Berg Balance Scale
CBD: corticobasal degeneration
GDS: Geriatric Depression Scale
IPAQ: International Physical Activity Questionnaire
LED: levodopa equivalent dose
MSA: multiple system atrophy
MET: metabolic equivalent of task
PD: Parkinson disease
PDQ-39: Parkinson Disease Questionnaire-39
PSP: progressive supranuclear palsy
QOL: quality of life
TUG: timed up-and-go
UPDRS: Unified Parkinson's Disease Rating Scale

Edited by L Buis; submitted 02.02.21; peer-reviewed by A Samuel, H Santos; comments to author 10.03.21; revised version received 17.03.21; accepted 02.07.21; published 31.08.21.

Please cite as:
Kim A, Yun SJ, Sung KS, Kim Y, Jo JY, Cho H, Park K, Oh BM, Seo HG
Exercise Management Using a Mobile App in Patients With Parkinsonism: Prospective, Open-Label, Single-Arm Pilot Study
JMIR Mhealth Uhealth 2021;9(8):e27662
URL: https://mhealth.jmir.org/2021/8/e27662
doi:10.2196/27662
PMID:34463635

Original Paper

Feasibility and Effectiveness of Assessing Subhealth Using a Mobile Health Management App (MibyeongBogam) in Early Middle-Aged Koreans: Randomized Controlled Trial

Younghwa Baek[1], PhD; Kyoungsik Jeong[1], BSc; Siwoo Lee[1], MD, PhD; Hoseok Kim[1], MSc; Bok-Nam Seo[1], MSc; Hee-Jeong Jin[1], PhD

Korean Medicine Data Division, Korea Institute of Oriental Medicine, Daejeon, Republic of Korea

Corresponding Author:
Hee-Jeong Jin, PhD
Korean Medicine Data Division
Korea Institute of Oriental Medicine
1672 Yuseongdae-ro
Yuseong-gu, Daejeon
Daejeon, 34054
Republic of Korea
Phone: 82 42 868 9305
Email: hjjin@kiom.re.kr

Abstract

Background: Mobile health (mHealth) is a major source of health management systems. Moreover, the demand for mHealth, which is in need of change due to the COVID-19 pandemic, is increasing worldwide. Accordingly, interest in health care in everyday life and the importance of mHealth are growing.

Objective: We developed the MibyeongBogam (MBBG) app that evaluates the user's subhealth status via a smartphone and provides a health management method based on that user's subhealth status for use in everyday life. Subhealth is defined as a state in which the capacity to recover to a healthy state is diminished, but without the presence of clinical disease. The objective of this study was to compare the awareness and status of subhealth after the use of the MBBG app between intervention and control groups, and to evaluate the app's practicality.

Methods: This study was a prospective, open-label, parallel group, randomized controlled trial. The study was conducted at two hospitals in Korea with 150 healthy people in their 30s and 40s, at a 1:1 allocation ratio. Participants visited the hospital three times as follows: preintervention, intermediate visit 6 weeks after the intervention, and final visit 12 weeks after the intervention. Key endpoints were measured at the first visit before the intervention and at 12 weeks after the intervention. The primary outcome was the awareness of subhealth, and the secondary outcomes were subhealth status, health-promoting behaviors, and motivation to engage in healthy behaviors.

Results: The primary outcome, subhealth awareness, tended to slightly increase for both groups after the uncompensated intervention, but there was no significant difference in the score between the two groups (intervention group: mean 23.69, SD 0.25 vs control group: mean 23.1, SD 0.25; *P*=.09). In the case of secondary outcomes, only some variables of the subhealth status showed significant differences between the two groups after the intervention, and the intervention group showed an improvement in the total scores of subhealth (*P*=.03), sleep disturbance (*P*=.02), depression (*P*=.003), anger (*P*=.01), and anxiety symptoms (*P*=.009) compared with the control group.

Conclusions: In this study, the MBBG app showed potential for improving the health, especially with regard to sleep disturbance and depression, of individuals without particular health problems. However, the effects of the app on subhealth awareness and health-promoting behaviors were not clearly evaluated. Therefore, further studies to assess improvements in health after the use of personalized health management programs provided by the MBBG app are needed. The MBBG app may be useful for members of the general public, who are not diagnosed with a disease but are unable to lead an optimal daily life due to discomfort, to seek strategies that can improve their health.

Trial Registration: Clinical Research Information Service KCT0003488; https://cris.nih.go.kr/cris/search/search_result_st01.jsp?seq=14379

(JMIR Mhealth Uhealth 2021;9(8):e27455) doi:10.2196/27455

KEYWORDS

mobile health; health status; mobile app; middle-aged group; subhealth; Korean medicine

Introduction

Mobile health (mHealth) using smartphone-based apps is poised to become a major source of health guidance. The "new normal" phenomenon induced by the COVID-19 pandemic is expected to further accelerate the digital economy. In health care, the representative keyword of the post-COVID-19 era is "digital (mobile) health care," which has become a necessity. Before the COVID-19 pandemic, the main targets of health care services were existing patients and older adults. However, the COVID-19 pandemic has increased the possibility that even healthy individuals can become patients, and this has increased the demand for health care services [1-3].

The World Health Organization stated that "the use of mobile and wireless technologies to support the achievement of health objectives has the potential to transform the face of health service delivery across the globe" [4], and mHealth is already used in various areas of health care. Statista predicted that the mobile health care market would continue to grow and that the total market value for mHealth applications in the US would exceed US $50 billion in 2025, which is approximately 25 times greater than the US $2 billion value in 2016 [5]. One US survey of "app users" showed that 31% of mobile phone owners used their phones to access health information, with the largest proportion (52%) being smartphone users [6].

mHealth is being developed for the management of not only daily healthy lifestyles, including aspects such as activity level, diet [7], and smoking cessation [8], but also chronic diseases, including hypertension [9] and diabetes [10], specific diseases, including juvenile idiopathic arthritis [11] and relapsed and refractory multiple myeloma [12], and physical and emotional aspects, such as pain [13], sleep [14], and depression [15]. In recent studies by Kitt et al, mHealth was found to be effective in reducing health care costs and improving health outcomes [9,16]. It is thought that mHealth contributes to continuous and active monitoring of health at individual or group levels [6], reduces and prevents health problems through promotion of health behaviors, supports self-management of chronic diseases, and improves the knowledge of health information, which can lead to fewer visits to medical institutions and a direct reduction of medical costs [6,17,18].

Traditional East Asian medicine (TEAM), which is mainly used in China, Korea, and Japan, was included in the "Supplementary Chapter Traditional Medicine Conditions—Module I" of the 11th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-11) in 2019. This means that TEAM is now officially recognized as a part of mainstream medical practice [19]. TEAM emphasizes preventative health management before the onset of diseases and focuses on subhealth management between disease and health. The term may differ in different countries; it is termed subhealth or *mibyeong* in traditional Korean medicine (TKM) [20]. *Mibyeong* is defined as a "state of discomfort in daily life due to abnormal symptoms, or abnormal examination findings, without a diagnosis of any disease, and as a result, a decrease in capacity to recover to a healthy state" [21]. The abnormal symptoms in *mibyeong* include fatigue, pain, sleep disturbance, and digestive disturbance, as well as emotional symptoms including depression, anger, and anxiety, which are the most common reasons for people to visit clinics or health care centers [21]. Nonetheless, conventional or physiological pathology does not clearly explain why some people have a *mibyeong* status, which may carry with it a high risk for future disease development [21]. Therefore, individuals with *mibyeong* must be aware of their health status and prioritize actively managing their own health. In this study, we developed a mobile app called MibyeongBogam (MBBG), which can be accessed on a smartphone to recognize and evaluate individual subhealth status and provide individualized health management strategies based on Korean medicine [22].

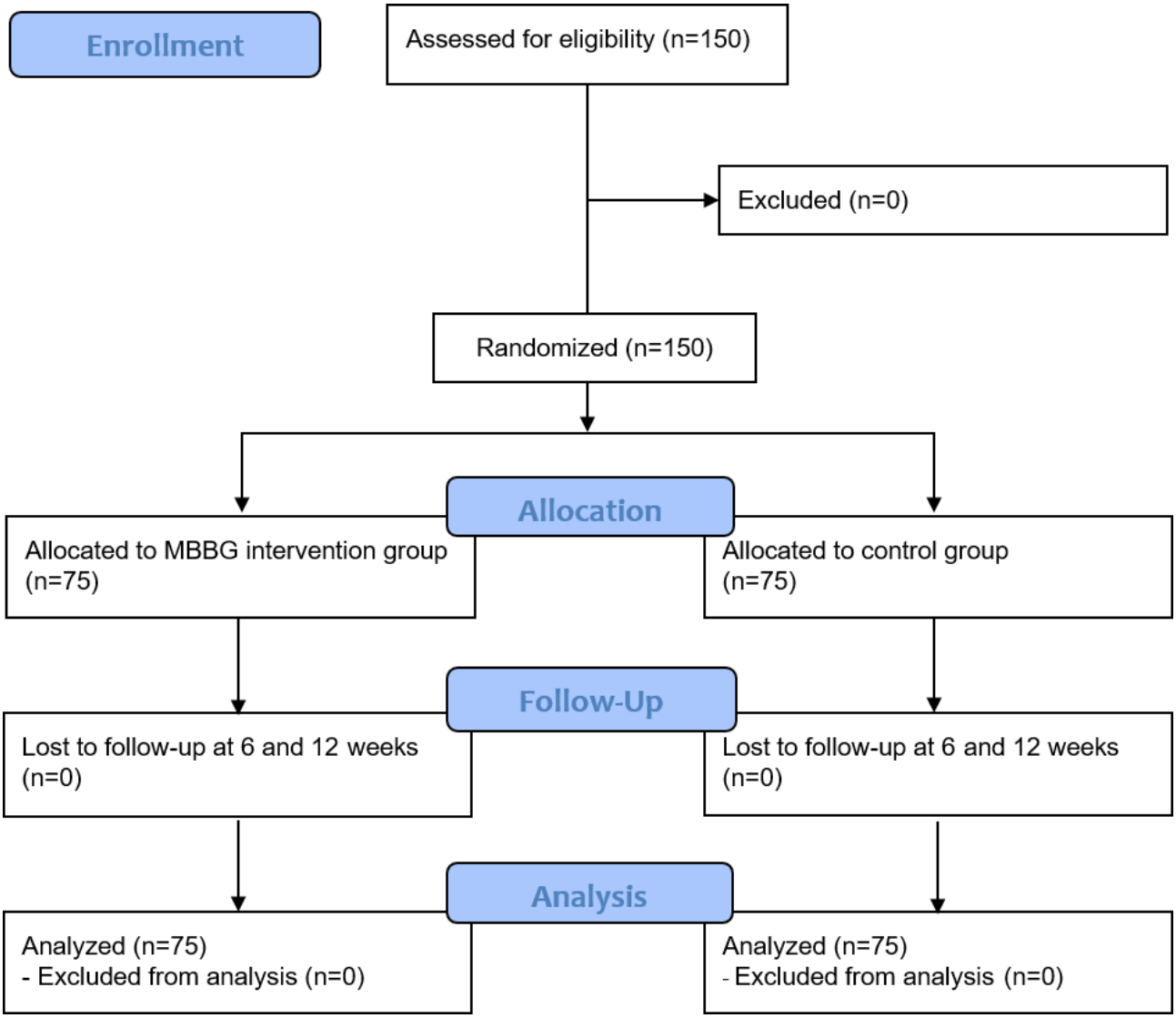
The objective of this study was to assess and compare the awareness of subhealth, changes in the subjective health status, and health behaviors between intervention (MBBG use group) and control groups. Based on these results, the feasibility of the MBBG app in managing and preventing a subhealth status in individuals was assessed.

Methods

Study Design

This study was a prospective, open-label, parallel group, randomized controlled trial. The protocol of this study has been described in detail in a previous study [23]. Selected participants visited the hospital three times, including before the intervention, at the 6-week posttest follow-up (first follow-up), and at the 12-week posttest follow-up (second follow-up, end of the intervention), and the main outcome variables were measured at the first visit before the intervention and at 12 weeks after the intervention (Figure 1). This study was conducted from November 2018 to February 2019 in two hospitals (Kyung Hee University Korean Medicine Hospital in Hoegidont, Seoul, and Kyung Hee University Korean Medicine Hospital in Gangdong, Seoul) on a total of 150 healthy participants in their 30s and 40s without any particular health problems. The eligible participants were randomly allocated to either the MBBG or control group, at a 1:1 allocation ratio. The MBBG group used the app for a total of 12 weeks, while the control group received no intervention.

Figure 1. CONSORT diagram of the study. MBBG: MibyeongBogam.



This study was approved by the Institutional Review Board at each institution (IRB numbers KOMCIRB-2018-07-002 and KHNMCOH 2018-07-002-001), and the physicians obtained written consent after all information regarding the study was provided to the participants. The protocol was registered in the Clinical Research Information Service (CRIS number KCT0003488). The anonymity and privacy of the participants were ensured as follows. Information regarding the collection and management of personal data (ie, phone number, email address, password, nickname, IP address, cookie content, etc) was provided, and consent was obtained when participants registered for the MBBG app in accordance with the Personal Information Protection Act. Moreover, app passwords were encrypted and stored in a database, and technical and physical protection measures against personal information leakage were established. Participants were also provided with personal IDs for the purpose of the study, to ensure anonymity.

Participants

Participants were recruited via posts on both online and offline boards and were screened. Healthy male and female adults, aged between 30 and 49 years, who were capable of using mobile smartphones, were eligible for this study. They were also required to complete self-report questionnaires and undergo physical examinations. If the participants did not own mobile smartphones with Android version 4.4 or higher or iOS version 9 or higher, they were excluded from the screening process. Any participants assessed and found to have clinically significant medical conditions through an interview with a physician, from their medical history (23 disease categories)/concomitant medication reviews and physical examinations, were also excluded from the study. If they were already using other mobile health care apps, they were ineligible. Participants who were involved in other trials in the preceding month of the study or were pregnant at baseline were also deemed ineligible.

Intervention

MBBG App

The intervention of this study was MBBG, a mobile app for subhealth management, developed by the Korea Institute of Oriental Medicine, Daejeon, Republic of Korea. MBBG aims to assess a user's subhealth status, as well as their TKM-based health status, based on which it recommends specific health-promoting strategies, such as meditation, exercise, and

consumption of herbal tea. Individuals can check their subhealth status and TKM health information after submitting all of the necessary information, including questionnaire responses. The questionnaires are included within the app so the participants can successfully complete these via the app. The physical examination results (height, weight, vital signs, pulse diagnosis, heart rate variability, etc) have to be inserted into the app manually or by automatic linkage [22] (Multimedia Appendix 1). In this study, all participants completed a survey questionnaire and underwent a physical examination on all three visits; only the MBBG group could access their results from the survey and physical examination by connecting with the MBBG app. The results of the control group, on the other hand, were uploaded to the MBBG app after the completion of the study.

MBBG App for the Intervention Group

After being allocated to the intervention group, participants first installed the MBBG app, after which they were educated on how to use the app verbally and with a user manual during each of their three visits. They were expected to use MBBG at least once daily for a total of 12 weeks. They accessed the app daily to read about their health status and ways to manage their health. In addition to hospital visits, the participants were free to complete the surveys and recommended health management protocols on the MBBG app, although this was not mandatory. The push notification function was activated to motivate and remind the participants to use the app throughout the study period. History tracking and user ranking services were also available to help promote the use of MBBG. Participants were not allowed to use any other mobile app for health management during the study period.

No Intervention in the Control Group

Participants allocated to the control group did not receive any intervention. They were told to maintain their usual lifestyle during the study period and were not allowed to use any mobile app for health management.

Outcome Assessment

Primary Outcome

The primary outcome was the awareness of subhealth, also known as *mibyeong* in TKM. The participants were given a questionnaire consisting of the following four items: (1) “Do you know or have you heard about subhealth status?” (2) “Do you think that preventing diseases is as important as treating them?” (3) “Do you think a professional medical service aimed at managing subhealth status is necessary?” and (4) “Are you willing to use a professional medical service to manage subhealth status, if available?” Each item was then scored from 1 (not at all) to 7 (absolutely), with the total score ranging from 4 to 28. All participants were required to submit the subhealth awareness questionnaire on their first and third visits. This questionnaire was independently developed in this study and was categorized into two factors (factor 1: item 1; factor 2: items 2, 3, and 4) based on a factor analysis. The Cronbach α of the questionnaire was .52 (the Cronbach α of factor 2, excluding item 1, was .82).

Secondary Outcomes

Secondary outcomes included subhealth status, health-promoting behaviors, and motivation for healthy behaviors. Subhealth status was evaluated using the *Mibyeong* questionnaire, which had a satisfactory reliability (Cronbach α=.88; intraclass correlation coefficient range: 0.67-0.83 in the test-retest method) and validity (correlation range: 0.47-0.48, compared to the SF-12, which is a well-known generic health status measure) [24,25]. The *Mibyeong* questionnaire consisted of 21 items on seven symptoms (fatigue, pain, sleep disturbance, digestive disturbance, depression, anger, and anxiety), and it assessed the severity, duration, and changes in those symptoms after rest in the preceding month. Each item was evaluated on a 7-point scale, and the total score ranged from 21 (healthy) to 147 (unhealthy). Higher scores indicate poor health status. The Cronbach α in this study was .88 (the Cronbach α ranged between .78 and .92 for individual symptoms).

Health-promoting behaviors are a measure of performance of health behaviors, which were evaluated using the Health Behavior Scale [26]. This scale consists of 25 items related to health responsibility (five items), diet habits (eight items), exercise (four items), stress management (five items), and smoking habits (three items). Each item has a 4-point response, from 1 (never) to 4 (always), with the total score ranging from 25 to 100. The higher the score, the more frequently the individual engages in healthy behaviors. The Cronbach α in this study was .78 (the Cronbach α ranged between .56 and .73 for individual domains).

In addition, the motivation for engaging in healthy behaviors is a measure of confidence in health behavior practice, and it was evaluated using the Self-Efficacy Questionnaire [27,28]. The six questions therein were on people’s abilities to avoid greasy food, quit smoking, exercise regularly, take necessary medications, relieve mental stress, and obtain health-related information. The responses were provided on a 4-point scale, where 1 is “not confident at all” and 4 is “absolutely confident.” The total score ranges from 6 to 24. Higher scores indicate higher confidence in behavior practice.

Feasibility Assessment

The feasibility of MBBG was assessed by evaluating the user finding access rate and the number of times participants logged onto the app during the intervention period. The user finding access rate was calculated using the number of times the app was accessed by the participants more than once a day, and the access rate for the 12-week intervention period was calculated.

Sample Size

The primary objective of this study was to compare the awareness levels of subhealth between the MBBG and control groups. Since there have not been any previous studies implementing the MBBG app, we conducted another clinical trial to explore the mental health benefits of a mobile app. In that trial, there was a 0.58 effect size with a 6-week test [29]. In this study, we set the intervention period as 12 weeks, and the age for study eligibility was higher; therefore, we assumed the effect size to be, conservatively, 0.5. Thus, the sample size was calculated as 60 per group (two-sided, α=.05, and

power=0.8, independent *t* test) using G-power software version 3.1.3. With a 20% dropout rate expected, we enrolled 150 participants (75 in each group).

Random Allocation Concealment and Blinding

Randomization was performed by a statistician prior to enrollment with an assignment ratio of 1:1 and a block size of 4. Information on participant group allocation was sealed in individual opaque envelopes that were consecutively numbered for allocation concealment. The investigators opened the envelopes in consecutive order and assigned the participants to either the MBBG or the control group after a screening assessment was conducted. Since this was an open-label study, the participants and investigators were not blinded. However, the outcome assessors were blinded throughout the study to minimize possible bias.

Statistical Analysis

In the preintervention survey, the Student *t* test and chi-square test were conducted to compare continuous and categorical variables, respectively, between the intervention and control groups. We performed an intention-to-treat analysis on all outcome measures, using the MBBG app at least once, and assessed the primary outcome at least once. Every participant participated in the study until the last day, and no values were excluded. An analysis of covariance (ANCOVA) was performed to compare the effects of the primary and secondary outcomes between the two groups after MBBG intervention. Age, sex, BMI, and the baseline value of each outcome variable were adjusted to calculate the least square means and standard errors. All analyses were performed using SAS 9.4 software (SAS Institute Inc), and statistical significance was set at $P<.05$.

Results

Study Population

A total of 150 participants were included in the study. Of these participants, 75 were randomly assigned to the MBBG group (23 men and 52 women) and 75 were assigned to the control group (21 men and 54 women). There were no differences in general characteristics, such as sex, age, and BMI, between the two groups, and the outcome variables were similar between the two groups, except for some specific variables, including total score of the subhealth status, pain, anger, and anxiety (Table 1).

Table 1. Participants' baseline demographics and outcome variable characteristics.

Variable	MBBG[a] group (n=75)	Control group (n=75)	*P* value
Sex (men/women), n	23/52	21/54	.86
Age (years), mean (SD)	41.73 (5.17)	42.09 (4.74)	.66
BMI (kg/m^2), mean (SD)	23.66 (3.61)	24.21 (4.06)	.38
Subhealth awareness score, mean (SD)	21.13 (2.85)	21.48 (3.08)	.48
Subhealth status (*mibyeong*) score, mean (SD)			
Total score	35.90 (14.20)	40.80 (12.40)	.03[b]
Fatigue	9.01 (4.17)	9.33 (3.16)	.60
Pain	4.77 (3.89)	6.31 (3.77)	.02[b]
Sleep disturbance	5.17 (4.11)	5.36 (3.35)	.76
Digestive disturbance	4.76 (3.27)	4.89 (1.96)	.76
Depression	4.55 (3.41)	4.91 (2.02)	.43
Anger	3.65 (2.26)	4.99 (2.27)	<.001[b]
Anxiety	4.03 (2.86)	5.00 (2.34)	.02[b]
Health-promoting behavior score, mean (SD)			
Total score	66.05 (8.40)	66.92 (0.97)	.53
Health responsibility	15.24 (2.74)	15.20 (2.38)	.93
Exercise	10.37 (2.50)	10.35 (2.52)	.95
Diet habits	17.36 (3.52)	18.35 (3.22)	.08
Stress management	13.07 (2.43)	13.00 (2.44)	.87
Smoking habits	10.01 (2.64)	10.03 (2.82)	.98
Motivation for healthy behaviors	18.85 (2.38)	19.14 (2.47)	.46

[a]MBBG: MibyeongBogam.

[b]$P<.05$.

Subhealth Effectiveness Assessment

Subhealth awareness, which is a primary outcome, tended to slightly increase for both groups after the MBBG intervention; however, there was no significant difference in the score between the two groups (MBBG group: mean 23.69, SD 0.25 vs control group: mean 23.1, SD 0.25; *P*=.09). For secondary outcomes, several variables of subhealth status showed significant differences between the two groups. In the MBBG group, subhealth total score, sleep disturbance, depression, anger, and anxiety improved compared to the findings in the control group (Table 2).

Table 2. Results of the subhealth effectiveness assessment using primary and secondary outcome measures at the 12-week follow-up after the intervention.

Variables	MBBG[a] group (n=75), least square mean (SE)	Control group (n=75), least square mean (SE)	*F* [b]	*P* value[b]
Subhealth awareness	23.69 (0.25)	23.1 (0.25)	2.94	.09
Subhealth status (mibyeong) score				
Total score	33.94 (1.10)	37.5 (1.10)	5.13	.03[c]
Fatigue	8.56 (0.37)	7.95 (0.37)	1.35	.25
Pain	5.02 (0.36)	5.57 (0.36)	1.12	.29
Sleep disturbance	4.52 (0.33)	5.57 (0.33)	5.18	.02[c]
Digestive disturbance	4.47 (0.25)	5.05 (0.25)	2.69	.10
Depression	3.85 (0.23)	4.85 (0.23)	9.07	.003[c]
Anger	3.51 (0.22)	4.36 (0.22)	6.79	.01[c]
Anxiety	3.68 (0.22)	4.51 (0.22)	6.93	.009[c]
Health-promoting behaviors				
Total score	68.27 (0.67)	68.47 (0.67)	0.04	.84
Health responsibility	15.83 (0.24)	15.80 (0.24)	0.05	.83
Exercise	10.39 (0.18)	10.46 (0.18)	0.08	.78
Diet habits	18.47 (0.26)	18.29 (0.26)	0.24	.62
Stress management	13.25 (0.21)	13.43 (0.21)	0.36	.55
Smoking habits	10.39 (0.18)	10.47 (0.18)	0.10	.75
Motivation for healthy behaviors	19.32 (0.20)	18.92 (0.20)	1.82	.18

[a]MBBG: MibyeongBogam app.

[b]ANCOVA analysis adjusted for sex, age, BMI, and the baseline value of each outcome variable.

[c]*P*<.05.

Feasibility Assessment

The retention rate was assessed by evaluating the user finding access rate of the MBBG app during the intervention period, and the retention rate was 75.1% (SD 15.9%, range 22%-100%) for the entire 12-week period. In particular, the mean access rate for the first 6 weeks postintervention was 71.9% (SD 17.7%, range 25%-100%), and the mean access rate for the next 6 weeks was 78.8% (SD 16.3%, range 14%-100%).

Discussion

Principal Results

This study is the first to compare changes in subhealth awareness and subhealth status after 12 weeks of using the MBBG app, which was developed as a framework based on the concept and management methods of TKM. This study also assessed the feasibility of the app as a self-guided preventative intervention. First, there was no significant difference in subhealth awareness between the MBBG and control groups; however, subhealth awareness tended to slightly increase in both groups. Second, the MBBG app showed positive effects on sleep, depression, anger, and anxiety, which are related to mental health. However, health-promoting behaviors and motivation for healthy behaviors were not significantly improved. This study is meaningful in that the MBBG app had significant effects on improving the health status in healthy adults, particularly the management of mental health symptoms.

Comparison With Prior Work

Awareness of Subhealth

In our study, there was no significant difference in the awareness of subhealth, which was a primary outcome, between the two groups. However, it tended to increase in both groups regardless of MBBG app usage. In our study, all participants in both groups met the researcher three times. The participants then received explanations on health and participated in health-related surveys. We suggest that processes, such as receiving explanations about

the study before consenting to participate, completing health-related questionnaires at each visit, and the health examination processes of measuring blood pressure and heart rate, would have partially contributed to the increased interest in participants' health awareness regardless of MBBG app usage. In a meta-analysis of mental health apps, using apps involving contact with medical staff was less effective than using apps without in-person feedback [30]. This is because a standalone app that does not promote contact with medical staff can enhance personal privacy and autonomy [31]. However, the main objective of our study was to assess changes in health awareness through the use of the MBBG app, that is, whether the participants became aware of the necessity of health care. Therefore, unlike the intervention effects of health apps observed in previous studies, it is thought that health-related information provided by medical staff, who were in contact with the participants, was an important factor of health awareness in our study. Additionally, previous studies that performed path analysis of cognitive factors related to the use of health apps demonstrated that the health consciousness of individual participants directly affected the use of health apps [32]. In our study, the mean pretest score of health awareness in the MBBG group was 21 out of 28, and a similar score was observed in the control group, suggesting that the participants in our study were already highly interested in health, which may be related to their health awareness.

Improvement of Mental Health

Interestingly, the subhealth status significantly improved in the MBBG group compared to the control group. Significant differences were observed in mental health aspects, such as sleep, depression, anger, and anxiety, between the two groups. These findings suggest that the MBBG app can improve mental health, especially discomfort, which is commonly observed in everyday life. The participants in our study belonged to the early middle-aged group, and these individuals often experience problems related to sleep, such as insufficient sleep time [33,34], decreased quality of sleep [35], and anxiety and depression symptoms [36]. Such symptoms are highly related to obesity, metabolic syndrome, and cardiovascular diseases [33,34,37,38]. However, most people do not seek or receive proper treatment for mental health problems. Recently, many scholars have predicted that technology-based interventions, such as health apps, have the potential to reduce treatment gaps in mental health. In addition, it is predicted that mental health apps will not replace the role of medical professionals in digital mental health and instead will play a role in interventions [39]. Moreover, a high level of evidence for the effects of smartphone-based interventions for common mental health problems, such as depressive symptoms, anxiety symptoms, stress levels, general psychiatric distress, quality of life, and positive effects, has been observed [40]. Approximately 41% of smartphone-based apps for mental health were developed for symptom relief, and these apps can help improve minor outcomes such as relaxation [41]. Furthermore, studies have reported that developing interest in mental health, acknowledging the problem, and undergoing interventions that can resolve minor symptoms at individual levels through health apps have positive effects on mental health in adults [40]. Therefore, the MBBG app developed in this study could serve as a health guide for those with physical and mental discomfort and those who cannot visit the appropriate hospital at the right time. A personalized health management strategy based on individual Korean medicine characteristics and discomfort is referred to as *Yangseng* in Korean medicine. This management strategy is further divided into herbal medicine, acupressure, exercise, and food in the MBBG app. Therefore, further studies on the positive effects of the MBBG app as an intervention in digital mental health care are required.

Change in Health Behaviors and Motivation

Health-promoting behaviors and motivation for health behaviors were not significantly different between the MBBG and control groups. Items on health responsibility (consultations with medical staff, health-related information acquisition, regular health examinations, etc), exercise (walking, high intensity exercise, etc), diet habits (regular meals, balanced food intake, etc), stress management (comfortable mindset, comfortable mindset, etc), and smoking habits (smoking cessation, overcoming the urge to smoke, etc) were used to assess the practice of and confidence in health-promoting behaviors. However, health-promoting behaviors and motivation did not significantly improve with MBBG app usage. A study by Ernsting et al focused on the use of health apps related to health-promoting behaviors such as smoking cessation, healthy diet, and weight loss. However, the authors argued that using health apps does not necessarily reflect the practice of health behaviors, but rather the motivation of users to change their health behaviors [42]. In addition, two systematic literature review studies reported different findings on the association between health apps and health-promoting behaviors. In the literature review of Lee et al on 12 studies that used health apps for health promotion programs, mobile app programs for the general public were mostly used for weight management and improvement of physical activities, and the effects of health-promoting behaviors were observed in those who used the apps for specific purposes compared to those who did not use the apps [43]. In contrast, in a study that reviewed 52 randomized controlled trials published between 2014 and 2019, there was no strong evidence to support the effects of mobile apps on improving health behaviors or outcomes [7]. Likewise, in this study, there was no significant difference between the MBBG and control groups. Therefore, it would be necessary to conduct a follow-up study by selecting appropriate participants and employing a detailed study design to assess the health-promoting effects of the MBBG app.

Lastly, the mean retention rate of the MBBG app in this study was 75.1%, which is similar to the rate of 79.6% (minimum 29%, maximum 100%) observed in a previous study [44]. In addition, the retention rate was defined as the number of initial study participants who remained in the study through the intervention period and follow-up in previous studies. In our study, the retention rate also included the daily app access rate of the participants, which reflected a high compliance. Similar results were observed in the dropout rate of participants. Although a dropout rate of 20% was predicted when designing the study and calculating the number of participants to include, the actual dropout rate was 0%. First, the participants of this

study were between the ages of 30 and 40 years and were comfortable or familiar with using mobile apps. A previous study reported that 44.3% of those aged between 30 and 40 years used health apps, which is higher than the proportion of app users in other age groups [45]. Second, this study was a feasibility study that assessed the use of the MBBG app and the change in awareness of subhealth. Thus, it is likely that the high degree of autonomy provided to the participants contributed to the low dropout rate.

Limitations

This study has several limitations. First, this study was conducted on participants in the early middle-age group. Therefore, generalization of the results to other age groups would be limited. However, this study is clinically and academically meaningful in that the feasibility of the app was evaluated in individuals in their 30s and 40s who required or needed to start taking more interest in health care. Second, the purpose of this study was to assess health status awareness and the feasibility of the MBBG app. Therefore, we could not assess whether the health management methods suggested by the MBBG app were implemented by the participants. Future studies should focus on the management strategies provided by the MBBG app and assess its effects. Third, only 150 participants were included in the study, and the 12-week intervention period was not long enough. However, the sample size in our study was similar to or slightly larger than that in other studies on mHealth interventions [29,30], and the intervention period was also similar to that in previous studies, which was 4-24 weeks [30]. Lastly, the participants and researchers were not blinded to randomization, which could have caused biased results. However, randomization was performed to control for adjusted variables, such as sex and age, which mainly affected the outcome variables between the two groups.

Conclusions

This randomized controlled trial compared the perception of and changes in the health status between intervention and control groups by using the MBBG app as an intervention for 3 months, and examined the possibility of using the MBBG app as a self-guided preventative intervention.

The MBBG app was developed to provide personalized health management strategies based on individual characteristics and self-awareness of the health status, which was assessed using symptoms, such as fatigue, sleep, and depression, which are commonly observed in daily life. In this study, the MBBG app did not significantly improve subhealth awareness. However, the MBBG app showed potential for improving health outcomes, especially in the mental health aspect, of individuals without particular health problems. We believe that the MBBG app would be useful for members of the general public, who are not diagnosed with a disease but do not enjoy optimal daily life due to discomfort, to seek strategies that can improve their health. Based on the feasibility of the app observed in this study, a large-scale randomized controlled trial would be necessary in the future. Detailed health status (eg, symptom types such as sleep disturbance and depression), specific health-promoting behaviors, and strategies to stimulate motivation based on user convenience are needed to evaluate the effects of the MBBG app. However, expansion of the contents of the MBBG app and development of customized health care guidelines should be prioritized before conducting a large-scale study.

Acknowledgments

This study was supported by the “Development of Korean Medicine Original Technology for Preventive Treatment based on Integrative Big Data” grant from the Korea Institution of Oriental Medicine (number KSN2022120) and by grants from the Ministry of Science, ICT & Future Planning (number NRF-2014M3A9D7034351).

Authors' Contributions

Conceptualization and methodology: YB, HJJ, and SL; data curation: BNS and KJ; statistical analysis: YB, KJ, and HK; writing-original draft preparation: YB and HJJ; final approval of the manuscript: all authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
MibyeongBogam app content.
[DOCX File , 446 KB - mhealth_v9i8e27455_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2883 KB - mhealth_v9i8e27455_app2.pdf]

References

1. van der Velden RM, Hermans AN, Pluymaekers NA, Gawalko M, Vorstermans B, Martens H, et al. Coordination of a remote mHealth infrastructure for atrial fibrillation management during COVID-19 and beyond: TeleCheck-AF. International Journal of Care Coordination 2020 Aug 27;23(2-3):65-70. [doi: 10.1177/2053434520954619]

2. Linz D, van der Velden RM, Gawalko M, Hermans AN, Pluymaekers NA, Hendriks JM. Remote management and education in patients with cardiovascular conditions during COVID-19 and beyond. Int J Cardiol Heart Vasc 2020 Oct;30:100646 [FREE Full text] [doi: 10.1016/j.ijcha.2020.100646] [Medline: 32984496]
3. Camacho-Rivera M, Islam JY, Rivera A, Vidot DC. Attitudes Toward Using COVID-19 mHealth Tools Among Adults With Chronic Health Conditions: Secondary Data Analysis of the COVID-19 Impact Survey. JMIR Mhealth Uhealth 2020 Dec 17;8(12):e24693 [FREE Full text] [doi: 10.2196/24693] [Medline: 33301415]
4. mHealth: New horizons for health through mobile technologies. World Health Organization. 2011. URL: https://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2021-08-05]
5. Statista. URL: https://www.statista.com [accessed 2021-01-26]
6. Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. P T 2014 May;39(5):356-364 [FREE Full text] [Medline: 24883008]
7. Milne-Ives M, Lam C, De Cock C, Van Velthoven MH, Meinert E. Mobile Apps for Health Behavior Change in Physical Activity, Diet, Drug and Alcohol Use, and Mental Health: Systematic Review. JMIR Mhealth Uhealth 2020 Mar 18;8(3):e17046 [FREE Full text] [doi: 10.2196/17046] [Medline: 32186518]
8. Vinci C, Brandon KO, Kleinjan M, Hernandez LM, Sawyer LE, Haneke J, et al. Augmented Reality for Smoking Cessation: Development and Usability Study. JMIR Mhealth Uhealth 2020 Dec 31;8(12):e21643 [FREE Full text] [doi: 10.2196/21643] [Medline: 33382377]
9. Jamaladin H, van de Belt TH, Luijpers LC, de Graaff FR, Bredie SJ, Roeleveld N, et al. Mobile Apps for Blood Pressure Monitoring: Systematic Search in App Stores and Content Analysis. JMIR Mhealth Uhealth 2018 Nov 14;6(11):e187 [FREE Full text] [doi: 10.2196/mhealth.9888] [Medline: 30429116]
10. Seely EW, Weitzman PF, Cortes D, Romero Vicente S, Levkoff SE. Development and Feasibility of an App to Decrease Risk Factors for Type 2 Diabetes in Hispanic Women With Recent Gestational Diabetes (Hola Bebé, Adiós Diabetes): Pilot Pre-Post Study. JMIR Form Res 2020 Dec 31;4(12):e19677 [FREE Full text] [doi: 10.2196/19677] [Medline: 33382039]
11. Butler S, Sculley D, Santos DS, Fellas A, Gironès X, Singh-Grewal D, et al. Usability of eHealth and Mobile Health Interventions by Young People Living With Juvenile Idiopathic Arthritis: Systematic Review. JMIR Pediatr Parent 2020 Dec 01;3(2):e15833 [FREE Full text] [doi: 10.2196/15833] [Medline: 33258786]
12. Biran N, Anthony Kouyaté R, Yucel E, McGovern GE, Schoenthaler AM, Durling OG, et al. Adaptation and Evaluation of a Symptom-Monitoring Digital Health Intervention for Patients With Relapsed and Refractory Multiple Myeloma: Pilot Mixed-Methods Implementation Study. JMIR Form Res 2020 Nov 17;4(11):e18982 [FREE Full text] [doi: 10.2196/18982] [Medline: 33200997]
13. Johnson A, Yang F, Gollarahalli S, Banerjee T, Abrams D, Jonassaint J, et al. Use of Mobile Health Apps and Wearable Technology to Assess Changes and Predict Pain During Treatment of Acute Pain in Sickle Cell Disease: Feasibility Study. JMIR Mhealth Uhealth 2019 Dec 02;7(12):e13671 [FREE Full text] [doi: 10.2196/13671] [Medline: 31789599]
14. Guillodo E, Lemey C, Simonnet M, Walter M, Baca-García E, Masetti V, HUGOPSY Network, et al. Clinical Applications of Mobile Health Wearable-Based Sleep Monitoring: Systematic Review. JMIR Mhealth Uhealth 2020 Apr 01;8(4):e10733 [FREE Full text] [doi: 10.2196/10733] [Medline: 32234707]
15. Aguilera A, Figueroa CA, Hernandez-Ramos R, Sarkar U, Cemballi A, Gomez-Pathak L, et al. mHealth app using machine learning to increase physical activity in diabetes and depression: clinical trial protocol for the DIAMANTE Study. BMJ Open 2020 Aug 20;10(8):e034723 [FREE Full text] [doi: 10.1136/bmjopen-2019-034723] [Medline: 32819981]
16. Kitt J, Fox R, Tucker KL, McManus RJ. New Approaches in Hypertension Management: a Review of Current and Developing Technologies and Their Potential Impact on Hypertension Care. Curr Hypertens Rep 2019 Apr 25;21(6):44 [FREE Full text] [doi: 10.1007/s11906-019-0949-4] [Medline: 31025117]
17. Aceto G, Persico V, Pescapé A. Industry 4.0 and Health: Internet of Things, Big Data, and Cloud Computing for Healthcare 4.0. Journal of Industrial Information Integration 2020 Jun;18:100129. [doi: 10.1016/j.jii.2020.100129]
18. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The Impact of mHealth Interventions: Systematic Review of Systematic Reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
19. World Health Assembly Update, 25 May 2019. World Health Organization. 2019. URL: https://www.who.int/news/item/25-05-2019-world-health-assembly-update [accessed 2021-01-26]
20. Lee J, Dong SO, Lee Y, Kim S, Lee S. Recognition of and interventions for (subhealth) in South Korea: a national web-based survey of Korean medicine practitioners. Integr Med Res 2014 Jun;3(2):60-66 [FREE Full text] [doi: 10.1016/j.imr.2013.09.001] [Medline: 28664079]
21. Lee J, Kim S, Lee Y, Song S, Kim Y, Lee S. The concept of Mibyeong (sub-health) in Korea: A Delphi study. European Journal of Integrative Medicine 2013 Dec;5(6):514-518. [doi: 10.1016/j.eujim.2013.07.010]
22. Lee YS, Jin HJ, Park DI, Lee SW. Development of Mobile Application for Preventive Management based on Korean Medicine: Mibyeongbogam. Journal of Sasang Constitutional Medicine 2018;30(1):66-73. [doi: 10.7730/JSCM.2018.30.1.66]
23. Shin S, Oh H, Kang M, Park M, Koh BH, Hwang M. Feasibility and effectiveness assessment of a mobile application for subhealth management: Study protocol for a randomized controlled trial. Medicine (Baltimore) 2019 May;98(21):e15704-e15708 [FREE Full text] [doi: 10.1097/MD.0000000000015704] [Medline: 31124947]

24. Lee Y, Baek Y, Park K, Jin H, Lee S. Development and validation of an instrument to measure the health status of healthy but unsatisfied people : Mibyeong index. Journal of Society of Preventive Korean Medicine 2016;20(3):45-53 [FREE Full text]
25. Jang E, Yoon J, Lee Y. The Evaluation of Reliability and Validity for Mibyeong Questionnaire. The Journal of the Society of Korean Medicine Diagnostics 2017;21(1):13-25 [FREE Full text]
26. Song R, June KJ, Kim CG, Jeon MY. Comparisons of motivation, health behaviors, and functional status among elders in residential homes in Korea. Public Health Nurs 2004 Jul;21(4):361-371. [doi: 10.1111/j.0737-1209.2004.21410.x] [Medline: 15260842]
27. Pender NJ, Walker SN, Sechrist KR, Frank-Stromborg M. Predicting Health-Promoting Lifestyles in the Workplace. Nursing Research 1990;39(6):326-332. [doi: 10.1097/00006199-199011000-00002]
28. Song R. Analyzing Motivational Factors to Predict Health Behaviors among Older Adults. Korean Journal of Adult Nursing 2006;18(4):523-532 [FREE Full text]
29. Reid SC, Kauer SD, Hearps SJ, Crooke AH, Khor AS, Sanci LA, et al. A mobile phone application for the assessment and management of youth mental health problems in primary care: a randomised controlled trial. BMC Fam Pract 2011 Nov 29;12(1):131 [FREE Full text] [doi: 10.1186/1471-2296-12-131] [Medline: 22123031]
30. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct 21;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
31. Anton MT, Jones DJ. Adoption of Technology-Enhanced Treatments: Conceptual and Practical Considerations. Clin Psychol (New York) 2017 Sep;24(3):223-240 [FREE Full text] [doi: 10.1111/cpsp.12197] [Medline: 28966479]
32. Cho J, Park D, Lee HE. Cognitive factors of using health apps: systematic analysis of relationships among health consciousness, health information orientation, eHealth literacy, and health app use efficacy. J Med Internet Res 2014 May 09;16(5):e125 [FREE Full text] [doi: 10.2196/jmir.3283] [Medline: 24824062]
33. Cho K, Cho E, Hur J, Shin D. Association of Sleep Duration and Obesity According to Gender and Age in Korean Adults: Results from the Korea National Health and Nutrition Examination Survey 2007-2015. J Korean Med Sci 2018 Dec 31;33(53):e345 [FREE Full text] [doi: 10.3346/jkms.2018.33.e345] [Medline: 30595686]
34. Grandner MA, Schopfer EA, Sands-Lincoln M, Jackson N, Malhotra A. Relationship between sleep duration and body mass index depends on age. Obesity 2015 Nov 02;23(12):2491-2498. [doi: 10.1002/oby.21247]
35. Zhang H, Li Y, Mo H, Qiu D, Zhao J, Luo J, et al. A community-based cross-sectional study of sleep quality in middle-aged and older adults. Qual Life Res 2017 Apr 12;26(4):923-933. [doi: 10.1007/s11136-016-1408-1] [Medline: 27620727]
36. Bansal P, Chaudhary A, Soni R, Sharma S, Gupta V, Kaushal P. Depression and anxiety among middle-aged women: A community-based study. J Family Med Prim Care 2015;4(4):576-581 [FREE Full text] [doi: 10.4103/2249-4863.174297] [Medline: 26985419]
37. Rutledge T, Linke S, Krantz D, Johnson BD, Bittner V, Eastwood JA, et al. Comorbid Depression and Anxiety Symptoms as Predictors of Cardiovascular Events: Results From the NHLBI-Sponsored Women's Ischemia Syndrome Evaluation (WISE) Study. Psychosomatic Medicine 2009;71(9):958-964. [doi: 10.1097/psy.0b013e3181bd6062]
38. Okubo N, Matsuzaka M, Takahashi I, Sawada K, Sato S, Akimoto N, Hirosaki University Graduate School of Medicine. Relationship between self-reported sleep quality and metabolic syndrome in general population. BMC Public Health 2014 Jun 05;14(1):562 [FREE Full text] [doi: 10.1186/1471-2458-14-562] [Medline: 24903537]
39. Wasil AR, Weisz JR, DeRubeis RJ. Three questions to consider before developing a mental health app. World Psychiatry 2020 Jun 11;19(2):252-253 [FREE Full text] [doi: 10.1002/wps.20757] [Medline: 32394568]
40. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019 Oct 09;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]
41. Radovic A, Vona PL, Santostefano AM, Ciaravino S, Miller E, Stein BD. Smartphone Applications for Mental Health. Cyberpsychol Behav Soc Netw 2016 Jul;19(7):465-470 [FREE Full text] [doi: 10.1089/cyber.2015.0619] [Medline: 27428034]
42. Ernsting C, Dombrowski SU, Oedekoven M, O Sullivan JL, Kanzler M, Kuhlmey A, et al. Using Smartphones and Health Apps to Change and Manage Health Behaviors: A Population-Based Survey. J Med Internet Res 2017 Apr 05;19(4):e101 [FREE Full text] [doi: 10.2196/jmir.6838] [Medline: 28381394]
43. Lee M, Lee H, Kim Y, Kim J, Cho M, Jang J, et al. Mobile App-Based Health Promotion Programs: A Systematic Review of the Literature. Int J Environ Res Public Health 2018 Dec 13;15(12):2838 [FREE Full text] [doi: 10.3390/ijerph15122838] [Medline: 30551555]
44. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
45. Xie Z, Nacioglu A, Or C. Prevalence, Demographic Correlates, and Perceived Impacts of Mobile Health App Use Amongst Chinese Adults: Cross-Sectional Survey Study. JMIR Mhealth Uhealth 2018 Apr 26;6(4):e103 [FREE Full text] [doi: 10.2196/mhealth.9002] [Medline: 29699971]

Abbreviations

MBBG: MibyeongBogam
mHealth: mobile health
TEAM: traditional East Asian medicine
TKM: traditional Korean medicine

Edited by L Buis; submitted 26.01.21; peer-reviewed by E Jang, C Jacob; comments to author 14.04.21; revised version received 11.05.21; accepted 09.07.21; published 19.08.21.

Please cite as:
Baek Y, Jeong K, Lee S, Kim H, Seo BN, Jin HJ
Feasibility and Effectiveness of Assessing Subhealth Using a Mobile Health Management App (MibyeongBogam) in Early Middle-Aged Koreans: Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(8):e27455
URL: https://mhealth.jmir.org/2021/8/e27455/
doi:10.2196/27455
PMID:34420922

Original Paper

Digital Natives' Preferences on Mobile Artificial Intelligence Apps for Skin Cancer Diagnostics: Survey Study

Sarah Haggenmüller[1], MSc; Eva Krieghoff-Henning[1], PhD; Tanja Jutzi[1], PhD; Nicole Trapp[1]; Lennard Kiehl[1], BSc; Jochen Sven Utikal[2,3], MD; Sascha Fabian[4], PhD; Titus Josef Brinker[1], MD

[1]Digital Biomarkers for Oncology Group, National Center for Tumor Diseases, German Cancer Research Center, Heidelberg, Germany

[2]Department of Dermatology, Heidelberg University, Mannheim, Germany

[3]Skin Cancer Unit, German Cancer Research Center, Heidelberg, Germany

[4]Department of Economics, University of Applied Science Neu-Ulm, Neu-Ulm, Germany

Corresponding Author:
Titus Josef Brinker, MD
Digital Biomarkers for Oncology Group
National Center for Tumor Diseases
German Cancer Research Center
Im Neuenheimer Feld 280
Heidelberg, 69120
Germany
Phone: 49 6221 32 19 304
Email: titus.brinker@dkfz.de

Abstract

Background: Artificial intelligence (AI) has shown potential to improve diagnostics of various diseases, especially for early detection of skin cancer. Studies have yet to investigate the clear application of AI technology in clinical practice or determine the added value for younger user groups. Translation of AI-based diagnostic tools can only be successful if they are accepted by potential users. Young adults as digital natives may offer the greatest potential for successful implementation of AI into clinical practice, while at the same time, representing the future generation of skin cancer screening participants.

Objective: We conducted an anonymous online survey to examine how and to what extent individuals are willing to accept AI-based mobile apps for skin cancer diagnostics. We evaluated preferences and relative influences of concerns, with a focus on younger age groups.

Methods: We recruited participants below 35 years of age using three social media channels—Facebook, LinkedIn, and Xing. Descriptive analysis and statistical tests were performed to evaluate participants' attitudes toward mobile apps for skin examination. We integrated an adaptive choice-based conjoint to assess participants' preferences. We evaluated potential concerns using maximum difference scaling.

Results: We included 728 participants in the analysis. The majority of participants (66.5%, 484/728; 95% CI 0.631-0.699) expressed a positive attitude toward the use of AI-based apps. In particular, participants residing in big cities or small towns (P=.02) and individuals that were familiar with the use of health or fitness apps (P=.02) were significantly more open to mobile diagnostic systems. Hierarchical Bayes estimation of the preferences of participants with a positive attitude (n=484) revealed that the use of mobile apps as an assistance system was preferred. Participants ruled out app versions with an accuracy of ≤65%, apps using data storage without encryption, and systems that did not provide background information about the decision-making process. However, participants did not mind their data being used anonymously for research purposes, nor did they object to the inclusion of clinical patient information in the decision-making process. Maximum difference scaling analysis for the negative-minded participant group (n=244) showed that data security, insufficient trust in the app, and lack of personal interaction represented the dominant concerns with respect to app use.

Conclusions: The majority of potential future users below 35 years of age were ready to accept AI-based diagnostic solutions for early detection of skin cancer. However, for translation into clinical practice, the participants' demands for increased transparency and explainability of AI-based tools seem to be critical. Altogether, digital natives between 18 and 24 years and between 25 and 34 years of age expressed similar preferences and concerns when compared both to each other and to results obtained by previous studies that included other age groups.

(JMIR Mhealth Uhealth 2021;9(8):e22909) doi:10.2196/22909

KEYWORDS

artificial intelligence; skin cancer; skin cancer screening; diagnostics; digital natives; acceptance; concerns; preferences; online survey

Introduction

Deep learning algorithms for image classification play an ever-increasing role in medicine and oncology. Researchers strive to automate and improve the assessment of various diseases, skin cancer in particular, with artificial intelligence (AI) tools. In experimental settings, the performance of AI algorithms achieved accuracies that were on par or even exceeded the results obtained by experienced dermatologists [1-6]. Since early-stage detection of melanoma increases the chances of survival significantly [5], improved AI-based cancer diagnostics might reduce mortality as well as health care expenditure [9-13]. Consequently, an accurate distinction between skin cancer and noncancer through AI-based solutions is of great interest to support diagnosis [3,13,14].

Recent studies about the patient perspective showed that participants expected synergy effects between physical skin examination and the use of mobile apps [15]. The vast majority of participants in another study, with or without previous history of melanoma, had a positive opinion on the use of AI in dermatology, in particular, when used as an assistance system [16]. However, studies have not yet investigated the clear application of AI technology in clinical practice nor the added value for younger user groups.

Here, we report the results of a survey-based study designed to evaluate how and to what extent young adults would be willing to accept AI-based mobile apps for early detection of skin cancer. The general attitudes, preferences, and concerns of potential future skin cancer screening participants below 35 years of age were elaborated, as digital natives may offer the greatest potential for successful implementation of digital assistance systems in clinical practice [15].

Methods

Data Collection

We conducted an anonymous online survey using Sawtooth SSI Web Lighthouse Studio 9.8.1. Prior to gathering responses, to ensure comprehensibility and consistency, we tested the survey with 12 volunteers who had no professional background in AI. We then conducted the study with the online survey between March 18, 2020, and April 18, 2020. As we wanted to investigate the preferences and concerns of digital natives, the survey was advertised on three social media channels: Facebook, LinkedIn, and Xing. The survey language was German; the results were translated into English for this paper. Participation was voluntary, and anonymity was ensured by design, to increase the proportion of questions that were answered thoroughly and truthfully.

Participants' general outlook on using apps for skin cancer examination was collected by asking, "Can you generally imagine covering parts of your skin cancer examination with medical apps?" with response options of "definitely," "rather yes," "rather not," and "definitely not." Prior experience with health or fitness apps was collected by asking the yes or no question of "Do you use apps to track your health or vital signs?" We included additional questions to obtain sociodemographic data.

To obtain a detailed assessment of the preferences of participants who felt generally positive about mobile AI-based apps (n=484), an adaptive choice-based conjoint (ACBC) was integrated into the survey. Based on the insights from our preliminary qualitative research, 7 app features and corresponding level options were developed for this investigation (Multimedia Appendix 1). Moreover, to ensure that no unrealistic combinations were presented, certain prohibitions were specified (Multimedia Appendix 2).

The ACBC process typically consists of 3 parts. First, in the so-called build-your own section, participants created their own customized product and familiarized themselves with the relevant app features as well as the corresponding levels. In the next part—the screening section—apps with specific feature combinations were presented, and participants were asked whether they would consider using app versions with these combinations. Finally, within a choice tournament, participants were asked to choose their preferred product from a range of apps based on their answers in the previous parts.

For evaluation of frequently cited concerns about AI-based tools, a maximum difference scaling (MaxDiff) section was included in the survey with a special focus on participants that generally refused the use of mobile skin examination tools (n=244). Based on the insights from preliminary qualitative research as well as literature review [16,17], 6 inhibitory aspects were selected for the analysis. Participants were shown several subsets of possible concerns and were asked to specify which one they considered the most and the least important. Participants made choices rather than expressing their strength of concerns with a numerical or rating scale. In this way, greater discrimination could be achieved and a comparison of the relative impact of participants' concerns was possible. Moreover, to ensure that no relevant issues were left out, participants received the opportunity to express additional concerns in a free-text question.

Data Validation

A total of 1548 participants below 35 years of age took part in the survey. To ensure data quality for subsequent analysis, responses that did not fulfill our internal quality criteria were identified and eliminated from the data set. For this purpose, a multilevel data cleaning process was applied. We excluded participants who answered only part of the questionnaire (n=731), participants living outside of Germany (n=36), and participants under the age of 18 (n=21). Furthermore,

contradictions in the participants' answers were examined; as a result, we removed another 2 participants. To minimize the risk of including participants who did not consider the topic seriously or possibly interrupted the survey, participants that answered the survey extremely fast or slow were left out [18,19]. After deleting responses of participants that took less than 2 minutes (n=15) or more than 60 minutes (n=15), a validated data set of 728 participants remained.

Data Analysis

To evaluate participants' general attitudes toward mobile apps for early detection of skin cancer, a descriptive analysis was conducted. The categories "definitely" and "rather yes" were summarized as a positive attitude while "rather not" and "definitely not" were summarized as a negative attitude toward mobile apps for skin examination. Statistical analysis was performed using SPSS, version 25.0 (IBM Corporation). Chi-square tests were performed to outline associations between sociodemographic characteristics and selected items of the questionnaire. We therefore conducted prespecified subgroup analyses on gender, residence, type of insurance, and prior experience with health or fitness apps. In the results section, we report only significant differences with a significance level set to $P<.05$ for all analyses. We computed 95% CIs for the main results using the normal distribution approximation.

ACBC data were analyzed using hierarchical Bayes estimation. The results were expressed in terms of counts, importance values, and utilities [20-22]. Count analysis examined how often certain levels were defined as unacceptable or must-have criteria within the screening section of the ACBC [22]. To evaluate the relevance of an attribute within the choice process of participants, average importance values were calculated [20,21]. Thus, for each feature, the utility value of the level that was regarded as most useful minus the level that was considered least useful represented the utility range (X). Subsequently, all utility ranges were summed (Y), and the share of each feature was determined based on the equation:

$$\text{Feature importance (\%)} = (X/Y) \times 100.$$

We calculated 95% CIs, taking the average feature importance score ±1.96 × SE. SE was computed by taking the SD of the importance score divided by the square root of the sample size. Part-worth estimation was performed to determine which feature levels were preferred from the participants' point of view. Utility values are presented as zero-centered differences within each feature.

MaxDiff data were expressed as sample mean scores and then rescaled to probability scores that reflect the likelihood that a concern was selected as "most important" within MaxDiff. We calculated 95% CIs, taking the average rescaled probability score ±1.96 × SE.

Results

Baseline Characteristics of the Study Sample

The demographic characteristics of the study sample are shown in Table 1. The median age was 24 years and the age distribution was fairly symmetrical.

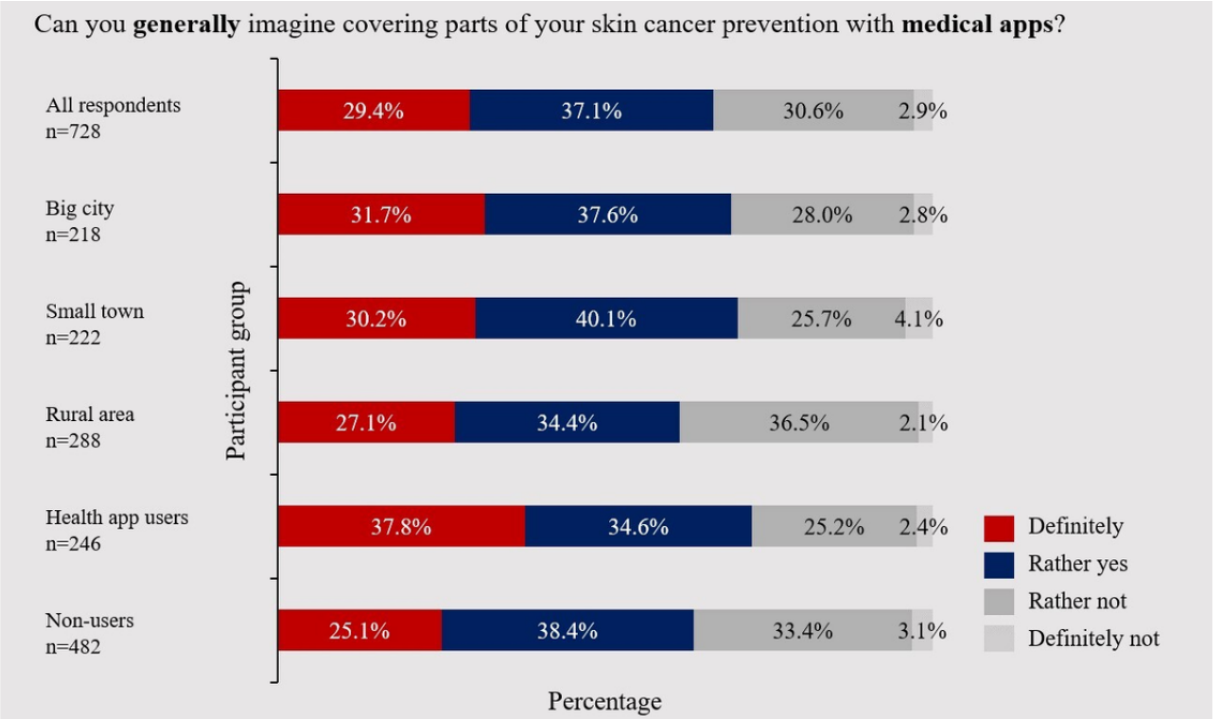
Table 1. Baseline characteristics of the study sample.

Sociodemographic characteristics	Values, n (%)
Gender	
Female	523 (71.8)
Male	205 (28.2)
Residence	
Large cities (>100,000 inhabitants)	218 (29.9)
Small towns (10,000-100,000 inhabitants)	222 (30.5)
Rural areas (<10,000 inhabitants)	288 (39.6)
Type of insurance	
Private insurance	82 (11.3)
Public insurance	646 (88.7)
Prior experience with health/fitness apps	
Yes	246 (33.8)
No	482 (66.2)
Age (years)	
18-24	420 (57.7)
25-34	308 (42.3)

General Attitude Toward Mobile Apps for Early Detection of Skin Cancer

Of all included participants (n=728), 484 participants (66.5%; 95% CI 0.631-0.699) were positive-minded toward the use of mobile apps for early detection of skin cancer. Only 21 participants explicitly ruled it out altogether (Figure 1). When comparing the age classes of 18 to 24 years and 25 to 34 years, no significant difference was observed (*P*=.97). Subgroup analysis revealed significant differences based on prespecified sociodemographic criteria. Out of the 440 participants residing in small towns or big cities, 307 (69.8%; 95% CI 0.655-0.740) felt significantly more positive toward mobile skin cancer screening apps than participants living in rural areas (177/288, 61.5%; 95% CI 0.558-0.671; *P*=.02). Moreover, previous experience with health or fitness apps had a significant effect on the willingness to use medical apps for early detection of skin cancer (*P*=.02).

Figure 1. Attitudes toward the use of mobile artificial intelligence–based apps for skin examination. Bar chart depicts the distribution of participants' general attitude depending on prespecified sociodemographic characteristics.



Of all participants that were generally negative-minded toward mobile apps for skin examination (n=244), 145 participants (59.4%; 95% CI 0.553-0.656) would rather consider using AI-based solutions if they received a reduced contribution from their health insurance company. In this context, participants between 18 and 24 years of age (94/141, 66.7%; 95% CI 0.589-0.745) were significantly more receptive to financial incentives than participants between 25 and 34 years of age (51/103, 49.5%; 95% CI 0.399-0.592; *P*=.007).

Preferences of the Future Generation of Skin Cancer Screening Participants

Must-Have and Unacceptable Criteria

For successful development of patient-usable AI systems, the identification of must-have and unacceptable criteria provides meaningful insights. The ACBC analysis identified app features that would cause future screening participants to reject the app. Of all participants that were generally positive-minded toward mobile apps for skin examination (n=484), 99 participants (20.5%; 95% CI 0.169-0.241) stated that they were not willing to rely on an exclusively app-based diagnosis. On the other hand, 179 participants of the 484 (37.0%; 95% CI 0.327-0.413), completely ruled out app versions with an accuracy of ≤65%. Data storage without encryption represented an exclusion criterion for 155 of the 484 participants (32.0%; 95% CI 0.279-0.362). Moreover, 100 of the 484 participants (20.7%; 95% CI 0.171-0.243) generally rejected apps without background information about the decision-making process. In line with this result, a further 96 participants of the 484 (19.8%; 95% CI 0.163-0.234) specified a basic explanation of the reasoning for the decision as an absolute must-have criterion.

Relative Importance of Individual App Features

Importance, on the other hand, indicates which relevance a certain app feature exerts on the decision-making process of participants [20,21]. Within this study, the app features "accuracy of the app," "field of application," and "data storage," on average, bore the greatest relative importance for participants' selection; the "accuracy of the app" constituted the top priority across all prespecified subgroups (Table 2). In contrast, "data processing" and "data usage" exerted only a minor influence on the choice of medical apps for skin cancer detection, thus opening up opportunities to pursue research interests without endangering the acceptance of future screening participants. Subgroup analysis on gender revealed that women attached considerably more importance to the app feature "explainability of the results" (females: 14.1%, males: 11.4%).

Table 2. Average relative importance of individual app features within ACBC.

Feature	Average importance, % (95% CI)
Accuracy of the app	20.4 (19.6-21.3)
Field of application	17.6 (16.9-18.3)
Data storage	17.4 (16.5-18.2)
Receipt of diagnosis	15.0 (14.2-15.7)
Explainability of the results	13.3 (12.7-14.0)
Data usage	9.8 (9.3-10.3)
Data processing	6.5 (6.2-6.8)

Preferred Levels of Individual App Features

After the relevance of individual app features within the process of choosing apps was examined, a detailed analysis was performed to determine which level of each app feature offers the greatest perceived usefulness from the participants' point of view [20] (overview of all app features and the corresponding levels, see Multimedia Appendix 1). This step is generally known as part-worth estimation.

Unsurprisingly, part-worth estimation revealed that the higher the accuracy level, the higher the added value from the participants' perspective. However, improving the accuracy from 85% to 90% disproportionately raised the perceived benefit (from 28.0 to 56.7), while an increase from 80% to 85% triggered only slight advancement in utility value (from 21.6 to 28.0). Consequently, achieving an accuracy of 80% might not lead to major shifts in the number of individuals willing to use a skin cancer screening app. The achievement of an accuracy of 90%, on the other hand, represented a convincing argument for the vast majority.

Concerning the feature "data usage," an integration of additional clinical information, such as age, gender, and patient medical history provided the greatest benefit for participants. Moreover, participants did not mind their data being used for research purposes in an anonymous way and, therefore, explicitly favored anonymous data processing for future research projects. These preferences were reflected regardless of sociodemographic characteristics across all prespecified subgroups.

Participants preferred an app scenario where the "field of application" is limited to appointment prioritization, followed by a personal consultation with a specialist. Moreover, participants favored that their health data be stored encrypted in the app and additionally protected by a personal password. Participants living in big cities would even prefer not to store their data at all, rather than storing them encrypted without personal password protection. The delivery of diagnostic results in real time constituted the preferred level regarding the receipt of diagnosis. The more detailed the explanation of the decision-making process, the higher the perceived usefulness was for participants.

Concerns of the Future Generation of Skin Cancer Screening Participants

Analyzing the MaxDiff data of this survey, no single criterion stood out significantly (Table 3). In terms of the probability scores, privacy concerns (24.3), insufficient trust in the app (23.5), and a lack of personal interaction (21.7) represented the dominant barriers from the participants' point of view. Moreover, concerns about incorrect app usage played a considerable role (17.7). In contrast, frequently mentioned aspects, such as the effort to deal with the functionality of the app (6.8) or the lack of technical affinity (6.0), exerted only minor influences.

Table 3. Evaluation of frequently cited concerns according to hierarchical Bayes estimation.

Ranking according to hierarchical Bayes estimation	Rescaled score[a] (95% CI)
Privacy concerns	24.3 (22.3-26.2)
Insufficient trust in the app	23.5 (22.1-25.0)
Lack of personal interaction	21.7 (20.0-23.5)
Incorrect app usage	17.7 (16.1-19.2)
Effort to deal with the functionality	6.8 (5.7-7.9)
Lack of technical affinity	6.0 (5.0-7.1)

[a]The rescaled score ranged from 0 to 100.

A total of 99 additions were made as part of the free-text entry. All responses were screened and analyzed in a qualitative manner. However, the majority of the answers had already been covered by the 6 selected main aspects, either by repeating or concretizing them, using examples. Beyond that, participants mentioned concerns regarding discrimination of elderly participants without smartphone experience, costs for both mobile devices and apps, and potential psychological burdens in case of suspected cancer.

Discussion

Principal Results

The majority of the future generation of skin cancer screening participants were ready to accept mobile apps for early detection of skin cancer. However, participants stated certain unacceptable and must-have criteria that need to be considered when developing patient-oriented AI-based solutions for dermatology. For translation into clinical practice, the demand for increased transparency and explainability appears particularly critical. Within the current state of AI, it is not possible to fully explain the reasoning of the decision making due to the black box phenomenon [23-26]. Therefore, to achieve broad acceptance among screening participants, approaches that encourage at least a basic explanation of the decision-making process are required.

The attributes “data processing” and “data usage” exerted only a minor influence on the choice of medical apps for skin examination; in fact, participants explicitly supported anonymous data processing and would therefore likely support the concept of open data, which encourages the sharing and release of data sets across research and clinical institutions. Moreover, from the participants’ point of view, there were no reservations against the inclusion of clinical information such as age, gender, and patient medical history.

For successful implementation into clinical practice, concerns of skeptical participants as well as identified rule-out criteria must be considered. Against this backdrop, the proper way to incorporate AI solutions within dermatology is by augmenting human intelligence and not replacing it. To leverage the potential of AI-based assistant systems, future research and clinical projects should emphasize personal interaction while simultaneously accomplishing a synergy between humans and AI systems. This approach coincides with the preferences of the majority of participants who were positive-minded toward the use of mobile apps for skin cancer detection in this ACBC, as well as with results obtained by previous studies [15,16]. Moreover, the active promotion of participants’ ability to act constitutes a key aspect. Mobile AI apps can only reach their full potential if future screening participants receive guidance and decision support. For individuals to trust Al-based apps, both orientation points and reliable and comprehensible health information are required. In this way, patients get an indication of how to distinguish potential medical AI assistance from conventional fitness or health apps. Furthermore, evaluation of participant concerns highlighted the demand for standardized regulations on how data are stored and protected within AI-based apps. This demand is also driven by the fact that 155 of 484 participants (32.0%) stated that data storage without encryption is an absolute exclusion criterion. Consequently, to achieve patient-oriented apps for dermatology, data security must play a key role within the whole development process.

Limitations

The baseline characteristics of this study sample showed that participants were predominantly female, thus not representative of the gender distribution in the general population. Since participants were recruited through social media, there is a risk of sampling bias, as social media users may be more likely to use mobile apps. Therefore, the results that we obtained are probably not fully generalizable to the general population of digital natives.

Importance was directly affected by the range of levels selected for each app feature as well as the total number of features [20,21]. Adding or removing a very popular or unpopular level to a feature would change the importance of all other attributes. Consequently, this paper could only reflect the importance relative to the features that were tested within this ACBC design.

MaxDiff, by definition, involves only comparative judgments. Thus, this elaboration cannot draw conclusions about the absolute magnitude of the selected impeding factors. One way to further increase the information value is to integrate additional questions that deliver an anchoring point (eg, specify the importance of one item), so that information in an absolute sense could be obtained [27].

Conclusions

The majority of potential screening participants below 35 years of age were ready to accept AI-based solutions. However, participants’ demands for increased transparency and explainability of AI-based tools must be considered for successful translation into clinical practice. Digital natives between 18 and 24 years and between 25 and 34 years of age showed similar preferences and concerns when compared to each other as well as to other age groups. They preferred the use of AI-based solutions as expert assistance systems, attached considerable value to the accuracy of AI apps, and expressed data privacy concerns.

Acknowledgments

This study was funded by the Federal Ministry of Health, Berlin, Germany (grants: Skin Classification Project 2 [SCP2] and Tumor Behavior Prediction Initiative [TPI], grant holder in both cases: Titus J. Brinker, MD, Digital Biomarkers for Oncology, National Center for Tumor Diseases, German Cancer Research Center, Germany). The sponsor had no role in the design and conduct of the study; collection, management, analysis and interpretation of the data; preparation, review or approval of the manuscript; and decision to submit the manuscript for publication.

Authors' Contributions

SH, SF, and TJB contributed to study concept and design. SH carried out the implementation, data collection, analysis, and statistics with contributions from SF, EKH, and TJ. SH, TJ, and EKH wrote the manuscript. TJB supervised the project. JSU, NT, and LK critically reviewed and edited the manuscript and approved the final version.

Conflicts of Interest

TJB would like to disclose that he is the owner of Smart Health Heidelberg GmbH (Handschuhsheimer Landstr. 9/1, 69120 Heidelberg, Germany), which develops mobile apps, outside of the submitted work. JSU is on the advisory board or has received honoraria and travel support from Amgen, Bristol Myers Squibb, GSK, LeoPharma, Merck Sharp and Dohme, Novartis, Pierre Fabre, Roche, and Sanofi, outside of the submitted work.

Multimedia Appendix 1
Overview about app features and the corresponding levels for this ACBC investigation.
[DOCX File , 15 KB - mhealth_v9i8e22909_app1.docx]

Multimedia Appendix 2
Prespecified prohibitions of this ACBC study.
[DOCX File , 13 KB - mhealth_v9i8e22909_app2.docx]

References

1. Maron RC, Weichenthal M, Utikal JS, Hekler A, Berking C, Hauschild A, et al. Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks. Eur J Cancer 2019 Sep;119:57-65. [doi: 10.1016/j.ejca.2019.06.013] [Medline: 31419752]
2. Haggenmüller S, Maron RC, Hekler A, Utikal JS, Barata C, Barnhill RL, et al. Skin Cancer Classification via Convolutional Neural Networks: Systematic Review of Studies Involving Human Experts. Eur J Cancer (forthcoming).
3. Han SS, Kim MS, Lim W, Park GH, Park I, Chang SE. Classification of the Clinical Images for Benign and Malignant Cutaneous Tumors Using a Deep Learning Algorithm. J Invest Dermatol 2018 Jul;138(7):1529-1538 [FREE Full text] [doi: 10.1016/j.jid.2018.01.028] [Medline: 29428356]
4. Brinker TJ, Hekler A, Enk AH, Berking C, Haferkamp S, Hauschild A, et al. Deep neural networks are superior to dermatologists in melanoma image classification. Eur J Cancer 2019 Sep;119:11-17 [FREE Full text] [doi: 10.1016/j.ejca.2019.05.023] [Medline: 31401469]
5. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017 Dec 02;542(7639):115-118. [doi: 10.1038/nature21056] [Medline: 28117445]
6. Brinker TJ, Hekler A, Enk AH, Klode J, Hauschild A, Berking C, Collaborators. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. Eur J Cancer 2019 May;113:47-54 [FREE Full text] [doi: 10.1016/j.ejca.2019.04.001] [Medline: 30981091]
7. Brinker TJ, Hekler A, Hauschild A, Berking C, Schilling B, Enk AH, et al. Comparing artificial intelligence algorithms to 157 German dermatologists: the melanoma classification benchmark. Eur J Cancer 2019 Apr;111:30-37 [FREE Full text] [doi: 10.1016/j.ejca.2018.12.016] [Medline: 30802784]
8. Brinker TJ, Hekler A, Enk AH, Klode J, Hauschild A, Berking C, Collaborators. A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task. Eur J Cancer 2019 Apr;111:148-154 [FREE Full text] [doi: 10.1016/j.ejca.2019.02.005] [Medline: 30852421]
9. Hoshyar AN, Al-Jumaily A, Hoshyar AN. The Beneficial Techniques in Preprocessing Step of Skin Cancer Detection System Comparing. Procedia Computer Science 2014;42:25-31. [doi: 10.1016/j.procs.2014.11.029]
10. Mehta P, Shah B. Review on Techniques and Steps of Computer Aided Skin Cancer Diagnosis. Procedia Computer Science 2016;85:309-316. [doi: 10.1016/j.procs.2016.05.238]
11. Premaladha J, Ravichandran KS. Novel Approaches for Diagnosing Melanoma Skin Lesions Through Supervised and Deep Learning Algorithms. J Med Syst 2016 Apr;40(4):96. [doi: 10.1007/s10916-016-0460-2] [Medline: 26872778]
12. Filho M, Ma Z, Tavares JMRS. A Review of the Quantification and Classification of Pigmented Skin Lesions: From Dedicated to Hand-Held Devices. J Med Syst 2015 Nov;39(11):177. [doi: 10.1007/s10916-015-0354-8] [Medline: 26411929]
13. Jain S, Jagtap V, Pise N. Computer Aided Melanoma Skin Cancer Detection Using Image Processing. Procedia Computer Science 2015;48:735-740. [doi: 10.1016/j.procs.2015.04.209]
14. Walker B, Rehg J, Kalra A, Winters R, Drews P, Dascalu J, et al. Dermoscopy diagnosis of cancerous lesions utilizing dual deep learning algorithms via visual and audio (sonification) outputs: Laboratory and prospective observational studies. EBioMedicine 2019 Feb;40:176-183. [doi: 10.1016/j.ebiom.2019.01.028]
15. Steeb T, Wessely A, Mastnik S, Brinker TJ, French LE, Niesert A, et al. Patient Attitudes and Their Awareness Towards Skin Cancer-Related Apps: Cross-Sectional Survey. JMIR Mhealth Uhealth 2019 Jul 02;7(7):e13844 [FREE Full text] [doi: 10.2196/13844] [Medline: 31267978]

16. Jutzi TB, Krieghoff-Henning EI, Holland-Letz T, Utikal JS, Hauschild A, Schadendorf D, et al. Artificial Intelligence in Skin Cancer Diagnostics: The Patients' Perspective. Front. Med 2020 Jun 2;7:233. [doi: 10.3389/fmed.2020.00233]
17. 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 Sep 4;6:255. [doi: 10.3389/fpubh.2018.00255]
18. Orme BK, Johnson RM. Testing Adaptive CBC: Shorter Questionnaires and BYO vs "Most Likelies". Research Paper Series Sawtooth Software Inc. 2008. URL: https://sawtoothsoftware.com/resources/technical-papers/testing-adaptive-cbc-shorter-questionnaires-and-byo-vs-most-likelies [accessed 2021-06-22]
19. Johnson RM, Orme BK. A New Approach to Adaptive CBC. In: Research Paper Series. 2007 Oct Presented at: Sawtooth Software Conference; October 17-19, 2007; Santa Rosa, California p. 85-109.
20. Steiner M, Meißner M. A User's Guide to the Galaxy of Conjoint Analysis and Compositional Preference Measurement. MAR 2018;40(2):3-25. [doi: 10.15358/0344-1369-2018-2-3]
21. Orme BK. Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research. Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research. Research Pub Llc; 2010.
22. Sawtooth S, Inc. ACBC Technical Paper. Technical Paper Series Sawtooth Software Inc. 2014. URL: https://www.sawtoothsoftware.com/download/techpap/acbctech2014.pdf [accessed 2021-06-22]
23. Chiu J, Wang Y, Su Y, Wei L, Liao J, Li Y. Artificial neural network to predict skeletal metastasis in patients with prostate cancer. J Med Syst 2009 Apr;33(2):91-100. [doi: 10.1007/s10916-008-9168-2] [Medline: 19397093]
24. Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. Lancet 2019 Apr 20;393(10181):1577-1579. [doi: 10.1016/S0140-6736(19)30037-6] [Medline: 31007185]
25. Wagholikar KB, Sundararajan V, Deshpande AW. Modeling paradigms for medical diagnostic decision support: a survey and future directions. J Med Syst 2012 Oct;36(5):3029-3049. [doi: 10.1007/s10916-011-9780-4] [Medline: 21964969]
26. Bach S, Binder A, Montavon G, Klauschen F, Müller KR, Samek W. On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. PLoS One 2015;10(7):e0130140 [FREE Full text] [doi: 10.1371/journal.pone.0130140] [Medline: 26161953]
27. Orme BK. How Good is Best Worst Scaling? And Is It More than Just Rank Order Data? Website Sawtooth Software Inc. 2018. URL: https://sawtoothsoftware.com/uploads/sawtoothsoftware/originals/How%20Good%20Is%20Best-Worst%20Scaling%20(2018).pdf [accessed 2021-06-22]

Abbreviations

ACBC: adaptive choice-based conjoint
AI: artificial intelligence
MaxDiff: maximum difference scaling

Edited by G Eysenbach; submitted 27.07.20; peer-reviewed by I Gabashvili, J Makin; comments to author 19.09.20; revised version received 02.10.20; accepted 13.04.21; published 27.08.21.

Please cite as:
Haggenmüller S, Krieghoff-Henning E, Jutzi T, Trapp N, Kiehl L, Utikal JS, Fabian S, Brinker TJ
Digital Natives' Preferences on Mobile Artificial Intelligence Apps for Skin Cancer Diagnostics: Survey Study
JMIR Mhealth Uhealth 2021;9(8):e22909
URL: https://mhealth.jmir.org/2021/8/e22909
doi:10.2196/22909
PMID:34448722

Viewpoint

Enhancing Healthcare Access–Smartphone Apps in Arrhythmia Screening: Viewpoint

Marcin Książczyk[1,2], MD; Agnieszka Dębska-Kozłowska[1], MD, PhD; Izabela Warchoł[1], MD; Andrzej Lubiński[1], MD, Prof Dr

[1]Department of Interventional Cardiology and Cardiac Arrhythmias, Medical University of Lodz, Łódź, Poland

[2]Department of Noninvasive Cardiology, Medical University of Lodz, Łódź, Poland

Corresponding Author:
Marcin Książczyk, MD
Department of Interventional Cardiology and Cardiac Arrhythmias
Medical University of Lodz
Ul. Żeromskiego 113
Łódź, 90-549
Poland
Phone: 48 42 639 35 63
Email: marcin_ksiazczyk@interia.pl

Abstract

Atrial fibrillation is the most commonly reported arrhythmia and, if undiagnosed or untreated, may lead to thromboembolic events. It is therefore desirable to provide screening to patients in order to detect atrial arrhythmias. Specific mobile apps and accessory devices, such as smartphones and smartwatches, may play a significant role in monitoring heart rhythm in populations at high risk of arrhythmia. These apps are becoming increasingly common among patients and professionals as a part of mobile health. The rapid development of mobile health solutions may revolutionize approaches to arrhythmia screening. In this viewpoint paper, we assess the availability of smartphone and smartwatch apps and evaluate their efficacy for monitoring heart rhythm and arrhythmia detection. The findings obtained so far suggest they are on the right track to improving the efficacy of early detection of atrial fibrillation, thus lowering the risk of stroke and reducing the economic burden placed on public health.

(JMIR Mhealth Uhealth 2021;9(8):e23425) doi:10.2196/23425

KEYWORDS

arrhythmia screening; atrial fibrillation; mobile electrocardiography; mobile health; phonocardiography; photoplethysmography; seismocardiography; stroke prevention

Introduction

The most commonly reported arrhythmia is atrial fibrillation (AF) [1]. Its prevalence is still underestimated [2], particularly the asymptomatic form: silent AF. Even so, the prevalence of symptomatic AF is estimated to be 0.12%-0.16% in patients aged <49 years, 3.7%-4.2% in patients aged 60-70 years, and almost 10%-17% in those aged ≥80 years [3]. The most common undiagnosed and untreated AF complications are thromboembolic events, such as stroke, which occur up to 5.6 times more frequently in AF patients [4]. It is therefore desirable to provide screening to patients in order to detect atrial arrhythmias. Additionally, the European Society of Cardiology (ESC) 2020 ESC Guidelines for the diagnosis and management of AF recommends opportunistic screening for AF by pulse taking or electrocardiogram (ECG) rhythm strip in patients above 65 years of age, and systematic ECG screening in patients above 75 years of age or those at high risk of stroke [5]. A problem arises when occasionally performed ECG does not record any arrhythmia, and the patient demonstrates palpitations or even worse symptoms, such as a thromboembolic event. As the prevalence of silent AF is estimated to be 10%-25% in the general population [6] and 30%-44% in older adults [7], it is reasonable to promote active screening for AF in patients at risk of the disease.

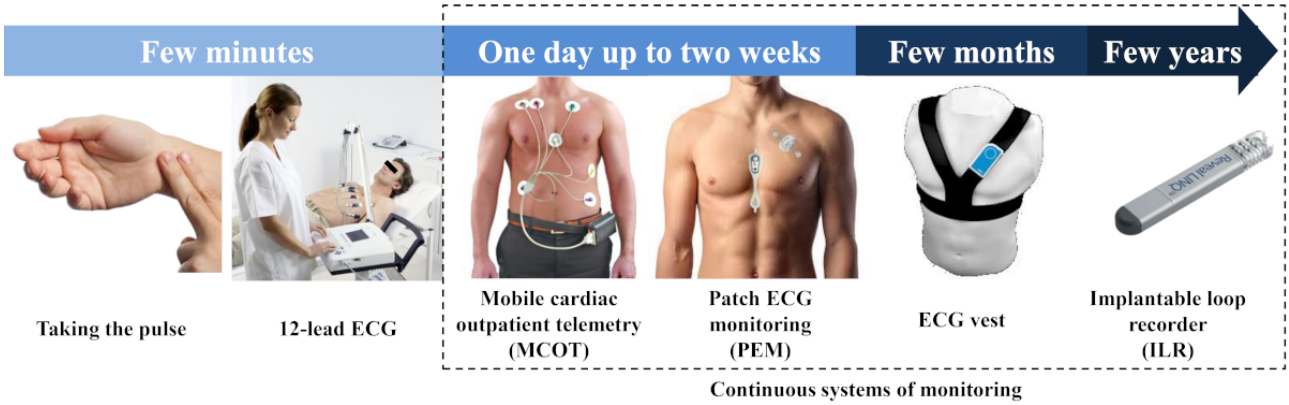
Specific mobile apps and accessory devices, such as smartphones and smartwatches, may play a significant role in monitoring heart rhythm in populations who are at high risk of arrhythmia: almost 2.71 billion smartphones are currently in use [8], and almost 150 million smartwatches are predicted to be in use in 2021 [9,10]. In general, the algorithms used by the apps correctly detect AF; however, if an automatic algorithm improperly classifies a trace as AF, it can then be verified and reclassified by a clinician.

In this viewpoint paper, we assess the availability of smartphone and smartwatch apps and evaluate their efficacy for monitoring heart rhythm and arrhythmia detection. These apps are becoming increasingly common among patients and professionals as a part of mobile health (mHealth) [11].

Methods of Screening for Arrhythmias and Heart Rhythm Monitoring

Practically, heart rhythm is typically monitored using continuous and intermittent systems. Continuous systems record the heart rhythm continuously from 24 hours up to 3 years; these show ECG varying in duration or with different numbers of presented leads (Figure 1).

Figure 1. Standard systems of heart rhythm monitoring [12-17]. ECG: electrocardiogram; ILR: implantable loop recorder; MCOT: mobile cardiac outpatient telemetry; PEM: patch electrocardiogram monitoring.



In contrast, intermittent systems are easily accessible and may play a role when continuous monitoring fails or is unacceptable by the patient. They record the heart rhythm on demand and are typically used upon symptom occurrence or according to a routine schedule (ie, each morning). For this paper, intermittent systems are classified into 5 main groups: standalone devices (ie, MyDiagnostick [18], The Heart Check PEN [19], or Lohman Afib Alert [20]), smartphone/smartwatch apps not dependent on an accessory device, smartphone apps dependent on an accessory device, and smartwatch apps.

The apps created for heart rhythm monitoring record the signal by either photoplethysmography (PPG), electrocardiography (ECG), seismocardiography (SCG), or phonocardiography (PCG). Of these, PPGs and ECGs have achieved commercial success. Some apps have been cleared by the United States Food and Drug Administration (FDA) or certified with the Conformité Européenne (CE) mark. They differ regarding their availability for particular mobile operating systems, duration of sample recording, and their ability to detect irregularity or even differentiate AF from normal sinus rhythm or other arrhythmias.

Methods for Identifying Available Apps

The European (Poland) App Store and Google Play were searched by 2 independent reviewers (MK and IW) for mobile apps that monitor heart rate. The search was performed between the September 9 and September 16, 2019. The apps offered in the App Store were searched using an iPhone 7 Plus with iOS 12.4.1 (Apple Inc), while those offered in Google Play were searched using a Samsung Galaxy S6 (Samsung Electronics) with Android Oreo 8.1 (Google). The following search string was employed: “heart rate” OR “atrial fibrillation” OR “ECG”. The inclusion criterion comprised the presence of an analogous or automatic algorithm for arrhythmia detection; no exclusion criteria were applied.

The Overview of Various Technologies and Apps

A total of 7 Android or iOS accessory device–independent smartphone or smartwatch apps, 8 Android or iOS smartphone accessory device–dependent apps, and 4 Android Wear/watchOS smartwatch apps were identified. An accessory device is defined as a tool with at least 2 built-in electrodes which wirelessly connects to a smartphone and is managed from a dedicated app. In addition to “core” apps that were identifiable in the search (Cardiio Pulsometer, Preventicus Heartbeats, and Kardia Mobile), 4 selected “mother-derived” apps were also evaluated: Cardiio Rhythm, Preventicus Nightwatch (both not available commercially), Kardia Band, and Kardia Mobile 6L. Information about these mother-derived apps are available on the developer's website. Due to the prevalence of ECG-based testing among the apps, 2 ECG-based smartphone apps (Kardia and Istel ECG) and 1 smartwatch app (Health) are presented as representative cases. All identified apps and their characteristics are presented in Tables 1 and 2 .

Table 1. Characteristics of apps used for heart rhythm monitoring.

App	Mobile operating system	Ratings	Downloads, n[a]	Cost (US $)	Method of recording
Accessory device–independent apps					
Heart_Rhythm	iOS	2.2/5.0	N/A[b]	Free at all	PPG[c]
Photo Afib Detector	iOS	2.0/5.0	N/A	Free at all	PPG
Cardiio: Pulsometr[d]	iOS	4.7/5.0	N/A	Free up to 9.99 per month	PPG
Cardiio Rhythm[e]	iOS	N/A	N/A	N/A	PPG
Preventicus Heartbeats	Android or iOS	4.1/5.0 4.3/5.0	100,000+	Free up to 43.99 per year	PPG
FibriCheck	Android or iOS	3.9/5.0 4.6/5.0	100,000+	4.71 up to 12.96 per month	PPG
Heart Beat	Android	3.5/5.0	10,000+	Free	PCG
BeatScanner	iOS	2.0/5.0	N/A	Free	SCG[f]
Accessory device–dependent apps					
Kardia Mobile[g]	Android or iOS	3.6/5.0 4.8/5.0	100,000+	Free[h]	ECG[i]
Kardia Mobile 6L	Android or iOS	3.6/5.0 4.8/5.0	100,000+	Free	ECG
Kardia Band for Apple Watch Series 1-3	iOS	3.6/5.0 4.8/5.0	100,000+	Free	ECG
ECG Check	Android or iOS	3.1/5.0 2.6/5.0	10,000+	Free	ECG
Istel ECG	Android or iOS	4.3/5.0 5.0/5.0	10,000+	Free	ECG
CardioSecur Pro	Android or iOS	3.4/5.0 4.3/5.0	5000+	Free	ECG
Sanket Life-ECG, Stress, Fitness	Android or iOS	3.2/5.0 3.3/5.0	1000+	Free	ECG
GEMS Mobile ECG for HeartCheck CardiBeat	Android or iOS	N/A 5.0/5.0	5000+	Free	ECG
Coala Heart Monitor	Android or iOS	4.0/5.0 3.9/5.0	1000+	Free	ECG
i2Dtx for CardioSleeve	iOS	5.0/5.0	N/A	Free	ECG
Smartwatch apps					
Preventicus Nightwatch[e]	Android Wear or watchOS	N/A	N/A	Free	PPG
FibriCheck	Fitbit OS	N/A	N/A	Free	PPG
Heart for Apple Watch: All series	watchOS	N/A	N/A	Free	PPG
ECG app for Apple Watch: Series 4 and subsequent	watchOS	N/A	N/A	Free	ECG
Huawei Health for Huawei Watch GT	Android Wear	N/A	N/A	Free	PPG
Heart Health for Garmin Watches	Android Wear	N/A	N/A	Free	PPG

[a]Data available only for Android apps.

[b]N/A: not applicable.

[c]PPG: photoplethysmography.

[d]Formerly known as Cardiio – Heart Rate.

[e]Not available commercially, study version only.
[f]SCG: seismocardiography.
[g]Formerly known as AliveCor.
[h]Device cost not included.
[i]ECG: electrocardiogram.

Table 2. Additional characteristics of apps used for heart rhythm monitoring.

App	Automatic irregularity or AF[a] detection algorithm	FDA[b] clearance	CE[c] certificate	Duration of recording	Number of leads if applicable
Accessory device–independent apps					
Heart_Rhythm	No	No	No	10 s	N/A[d]
Photo Afib Detector	Yes	No	No	30, 60, or 120 s	N/A
Cardiio: Pulsometr[e]	No	No	No	20 s	N/A
Cardiio Rhythm[f]	Yes	No	No	20 s	N/A
Preventicus Heartbeats	Yes	No	Ila	60 or 300 s	N/A
FibriCheck	Yes	Yes	Ila	60 s	N/A
Heart Beat	No	No	No	30 s	N/A
BeatScanner	Yes	No	No	120 s	N/A
Accessory device–dependent apps					
Kardia Mobile[g]	Yes	Yes	Ila	30 s	1
Kardia Mobile 6L	Yes	Yes	Ila	30 s	6
Kardia Band for Apple Watch Series 1-3	Yes	Yes	Ila	35 s	1
ECG Check	Yes	Yes	Ila	45 s	1
Istel ECG	Yes	No	Ila	30, 60, 120 or 180 s	6
CardioSecur Pro	Yes	No	Ila	30 s	6-12
Sanket Life-ECG, Stress, Fitness	No	No	Ila	20 s	1
GEMS Mobile ECG for HeartCheck CardiBeat	Yes	Yes	Ila	30-300 s	1
Coala Heart Monitor	Yes	Yes	Ila	60 s	2
i2Dtx for CardioSleeve	Yes	Yes	Ila	30 s	3
Smartwatch apps					
Preventicus Nightwatch[f]	Yes	No	Ila	Continuous	N/A
FibriCheck	Yes	Yes	Ila	60 s	N/A
Heart for Apple Watch: All series	No	No	Ila	Dependent on user activity	N/A
ECG app for Apple Watch: Series 4 and subsequent	Yes	Yes	Ila	30 s	1
Huawei Health for Huawei Watch GT	Yes	No	No	Dependent on user activity	N/A
Heart Health for Garmin Watches	Yes	No	No	Dependent on user activity	N/A

[a]AF: atrial fibrillation.

[b]FDA: Food and Drug Administration.

[c]CE: Conformité Européenne.

[d]N/A: not applicable.

[e]Formerly known as Cardiio – Heart Rate.

[f]Not available commercially, study version only.

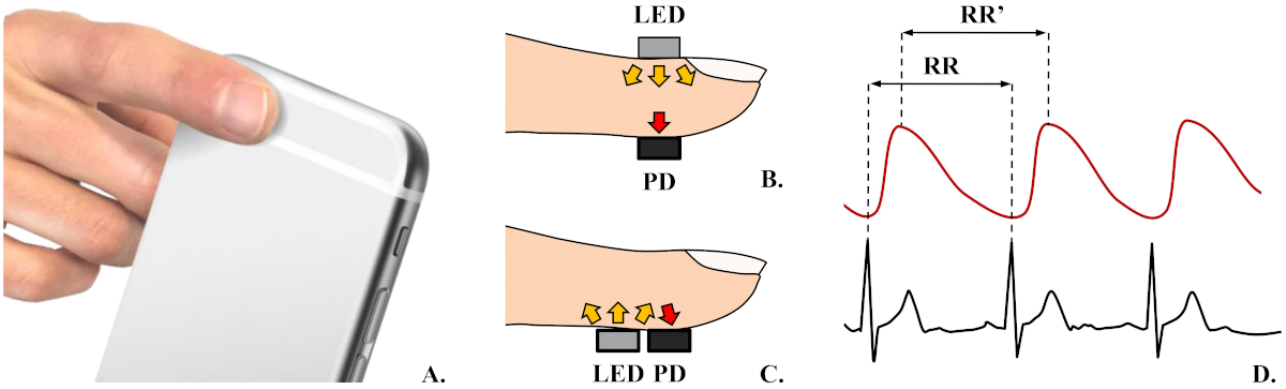
[g]Formerly known as AliveCor.

Apps Using PPG

PPG is a technology in which a light source, such as an light-emitting diode, illuminates a tissue, and a photodetector measures the amount of backscattered light returned [21]. The amount of backscattered light corresponds with the variations of blood volume over the sampling area. As blood volume is synchronous with heartbeat, PPG can accurately show heart

rate [22]. Nowadays, it is possible to obtain a photoplethysmogram in a patient suffering from cardiovascular disorders using a smartphone flash acting as a source of light and a camera serving as a photodetector (Figure 2A). Such photoplethysmograms are called "reflective", as both the light source and photodetector are on the same side of a fingertip. In contrast, systems where the light source and photodetector are located opposite to each other (Figure 2B and C), such as a pulse oximeter, are called "transmissive" [23]. A typical photoplethysmogram wave is shown in red in Figure 2D: its peaks are slightly delayed in relation to the R of the QRS complex in a standard electrocardiogram, representing the time the blood needs to fill up the furthest areas of the body.

Figure 2. (A) Measuring heart rate with mobile photoplethysmography. The finger is placed over the camera when the flash is on. (B) Transmissive method of measuring heart rate with PPG (used in pulse oximeters). (C) Reflective method of measuring heart rate with photoplethysmography (used in smartphones or smartwatches). (D) Differences in wave shape and RR to RR shift between photoplethysmogram (red curve) and electrocardiogram (black curve) [23-25]. LED: light-emitting diode; PD: photodetector; RR: the interval between 2 Rs in 1 heart cycle.



Heart_Rhythm

Heart_Rhythm is a free app that allows the user to record PPG and then compare the PPG with a model sinus rhythm or atrial fibrillation wave. The efficacy of such subjective self-assessments of rhythm patterns has not been validated in any clinical research [26].

Photo AFib Detector

Photo Afib Detector is a free app, which automatically detects an abnormality in the pattern of live-recorded PPG signal by estimating 2 statistical parameters: root mean square of successive difference and Shannon entropy [27]. An algorithm combining root mean square of successive difference and Shannon entropy in an iPhone 4S showed 96.2% sensitivity and 97.5% specificity for beat-to-beat discrimination of AF from sinus rhythm when compared with the 12-lead ECG [28]. However, Photo AFib Detector has not been directly validated in any clinical research.

Cardiio: Pulsometer (Former Name: Cardiio–Heart Rate Monitor) and Cardiio Rhythm

Cardiio: Pulsometer is a free app, while Cardiio Rhythm is a beta version currently used only for scientific purposes.

Cardiio: Pulsometer records high-quality PPG that can be evaluated by an expert and classified as sinus rhythm or rhythm other than sinus; unfortunately, there is no automatic algorithm for arrhythmia detection. Interestingly, the previous version of Cardiio: Pulsometer, called Cardiio – Heart Rate Monitor, was equipped with a face mode that enabled a contactless measurement of the heart rate based on the face of the user. Although Yan et al [29] showed that both finger and face PPGs demonstrate high accuracy in measuring resting heart rates, the app currently only uses the finger mode due to legal reasons [30].

Although Cardiio Rhythm is not currently commercially available, recent clinical findings regarding the app are promising. The sensitivity of the Cardiio Rhythm finger mode (92.9%, 95% CI 77-99) was found to be higher than the internet-enabled mobile ECG distributed by AliveCor (iECG; 71.4, 95% CI 51-87), while Cardiio Rhythm and iECG demonstrated comparable specificity (97.7%, 95% CI 97-99 vs 99.4%, 95% CI 99-100) [31]. Although Cardiio Rhythm demonstrated a lower positive predictive value (PPV) than did iECG (53.1%, 95% CI 38-67 vs 76.9%, 95% CI 56-91), both apps had high negative predictive values (NPV; 99.8%, 95% CI 99-100 vs 99.2, 95% CI 98-100) [31]. Cardiio Rhythm finger mode demonstrated 93.1% sensitivity (95% CI 86.9-97.2) and 90.9% specificity (95% CI 82.9-96.0) compared with superficial ECG, with a 92.2% PPV (95% CI 85.8-95.8) and 92.0% NPV (95% CI 94.8- 95.9) [32]. Finally, Cardiio Rhythm's facial mode effectiveness demonstrated high sensitivity (95%, 95% CI 87-98) and high specificity (96%, 95% CI 91-98) in discriminating AF compared with 12-lead ECG. The PPV and NPV of the facial mode was 92% (95 CI 84-96) and 97% (95% CI 93-99), respectively [33].

Preventicus Heartbeats and Preventicus Nightwatch

Preventicus Heartbeats is freely available for smartphones, while Preventicus Nightwatch is available only for smartwatch users. Both apps use PPG in screening for AF, and both have been validated in clinical trials.

The full version of Preventicus Heartbeats allows the user to record PPG and receive a complete report about the rhythm variability. In 2019, the Enhanced Diagnostics for Early Detection of Atrial Fibrillation–Prospective Validation (DETECT AF PRO) trial was performed to compare the efficacy of Preventicus Heartbeats in AF screening with iECG. The sensitivity and specificity of the Preventicus Heartbeats app increased with recording time from 1-3 to 5 minutes: the

sensitivity was found to be 89.9% (95% CI 85.5-93.4), 91.3% (95% CI 86.5-94.7), and 91.5% (95% CI 85.9-95.4), respectively, while the specificity was found to be 99.1% (95% CI 97.5-99.8), 98.7% (95% CI 96.7-99.6), and 99.6% (95% CI 97.8-100), respectively [34].

A similar trial, Smartwatches for Detection of Atrial Fibrillation (WATCH AF), was carried out to compare the efficacy of heart rhythm monitoring by the Preventicus Nightwatch smartwatch PPG-based algorithm with that of iECG. One-minute recordings were analyzed by the Preventicus Nightwatch (available for smartwatches only) and compared with the iECG. The algorithm demonstrated 93.7% sensitivity (95% CI 89.8-96.4) and 98.2% specificity (95% CI 95.8- 99.4) in detecting AF [35,36]. Preventicus Nightwatch appear to represent a breakthrough in the monitoring of arrhythmia as it will be able to continuously analyze PPG and document AF events lasting for at least 1 minute. However, it still remains in testing [35].

FibriCheck

Fibricheck is the only PPG-based heart rhythm monitoring app cleared both by the FDA and CE. In one study, a comparison of heart rate measurements by FibriCheck and 2 other FDA-cleared devices, Nonin oximeter and AliveCor, found a correlation of 0.834 between FibriCheck and Nonin, 0.88 between FibriCheck and AliveCor, and 0.897 between Nonin and AliveCor (no significant difference; *P*=.61); in addition, an R-R and peak-to-peak interval correlation of 0.993 was found between FibriCheck and wearable ECG (no significant difference; *P*=.92) [37].

FibriCheck was also included in the Real Life Digital Population Screening for Atrial Fibrillation Using only a Smartphone (DIGITAL AF II) study, including over 60,000 participants who completed the monitoring period. The study yielded a database of nearly 600,000 pieces of 1-minute PPGs [38]. Of these, 791 participants (1.3%) presented a trace typical for AF. The prevalence of AF in this population was found to be 1.68% in patients aged 40-49 years, 2.16% in those aged 50-59 years, 3.23% in those aged 60-69 years old, 5.97% in those aged 70-79 years, and 12.3% in those aged ≥80 years [38]. Unfortunately, the study has a few limitations: the traces were not compared with any other method, such as iECG or ECG, and only selected data were available. Elsewhere, FibriCheck demonstrated a sensitivity of 96% and a specificity of 91.1% compared with 12-lead ECG [39]. Its cost is not refundable from national health funds [40].

Heart for Apple Watch: All Series

The Heart app is an integral part of iOS and watchOS. All Apple Watch series use PPG to record the heart rate, but only series 4 and above are able to record ECG (see section ECG App for Apple Watch: Series 4 or Subsequent). However, Preventicus Nightwatch will be able to use a built-in algorithm to analyze the PPG traces recorded by Apple Watch to detect AF.

Apps Using Electrocardiography (Dependent on an Accessory Device)

Some mobile apps use ECG for recording and analyzing the signal and are dependent on accessory devices. These devices contain electrodes, whose number and location depend on whether 1-lead or 6-lead ECG is recorded. The devices examined in this paper are displayed in Figure 3, with the total number of the electrodes and recording leads shown in parentheses.

Figure 3. (A) Kardia Mobile (2 electrodes, 1-lead electrocardiogram). (B) Kardia Mobile 6L (3 electrodes, 6-lead electrocardiogram). (C) Kardia Band (2 electrodes, 1-lead electrocardiogram). (D) ECG Check (2 electrodes, 1-lead electrocardiogram). (E) Istel HR-2000 (4 electrodes, 6-lead electrocardiogram). (F) CardioSecur Pro (4 electrodes, 6-12-lead electrocardiogram). (G) Sanket Life-ECG, Stress, Fitness (3 electrodes, 1-12-lead electrocardiogram). (H) GEMS Mobile ECG for HeartCheck CardiBeat (2 electrodes, 1-lead electrocardiogram). (I) Coala Heart Monitor (3 electrodes, 2-lead electrocardiogram), J. i2Dtx for CardioSleeve (3 electrodes, 3-leads electrocardiogram). (K) Apple Watch Series 4 (3 electrodes, 1-lead electrocardiogram) [41-51].

Kardia Mobile

Kardia Mobile (former name: AliveCor) is a clinically validated mobile device for recording 1-lead ECG and the first to be cleared by the FDA [52]. The first of 2 studies that contributed to FDA clearance of iECG was conducted by Garabelli et al [53]. The obtained ECG curve corresponds to the first (I) limb lead. The Kardia Mobile app has a built-in automatic algorithm for arrhythmia detection focused on AF.

Although some kinds of arrhythmia, like premature ventricular/supraventricular contractions or conduction abnormalities (sinus bradycardia/tachycardia, bundle branch block, or atrioventricular block) may be improperly classified as AF or even unclassified by the automatic algorithm [54,55], the app has been updated to reduce the number of unclassified traces. A study on 214 patients found the single-channel ECG to demonstrate 90.9% sensitivity (95% CI 78.3-97.5) and 93.5% specificity (95% CI 88.7-96.7) for any rhythm abnormality, and 46.4% sensitivity (95% CI 27.5-66.1) and 100% specificity (95% CI 98.0-100) for any conduction abnormality [56]. As a result, even if an automatic algorithm improperly classifies the 1-lead ECG trace as AF, it may be correctly reclassified by a clinician.

A comparison of 1-lead ECG with lead I and II of 12-lead ECG in patients taking sotalol or dofetilide found reasonable agreement between measurements of corrected QT (QTc) interval in the sinus rhythm (bias 3 ms; SD of bias 46 ms) if QTc <500 ms [57].

The efficacy of Kardia Mobile in arrhythmia detection was validated in patients with cardiovascular implantable electronic devices. A study of recordings from 251 subjects with a pacemaker (59%) or implantable cardioverter-defibrillator (41%)

in paced and nonpaced states (if possible) found the readings to be adequately interpreted in 90% of paced recordings (25 of 251 recordings were "uninterpretable") and 94.7% of nonpaced recordings (9 of 171 recordings were "uninterpretable") [58].

Kardia Mobile is an effective tool for detecting arrhythmia or conduction abnormalities in children. It was found capable of detecting supraventricular tachycardia, AF, ectopic atrial tachycardia, atrial tachycardia, and ventricular tachycardia, and the users reported a high level of satisfaction [59]. In addition, a relationship was found between QRS dispersion and QTc intervals measured by 1-lead and 12-lead ECG in both healthy children and children with cardiac disease [60].

The QT intervals recorded by Kardia Mobile were 7 ms shorter than those from the 12-lead ECG, with only a 1.75% difference. In comparison, PQ intervals were found to be 20 ms shorter than those of conventional ECG, representing a more than 10% difference. Such a significant discrepancy between PQ intervals might lead to mimicking arrhythmias, otherwise known as pre-excitation syndrome [61].

Kardia Mobile 6L

The Kardia Mobile 6L is the first FDA-cleared 6-lead ECG. It has 3 built-in electrodes that record 6-lead ECG in channels I-III, aVR, aVL, and aVF [62]. The system uses the same app as the 1-lead Kardia Mobile. Thus far, it has not been included in clinical trials. It is expected that the new 6L will provide better-quality ECGs and greater information on ST-segment changes or axis determination than the standard AliveCor device.

Kardia Band

Kardia Band was the first FDA-cleared medical accessory for the Apple Watch Series 1 to 3 and replaced the original band. It has a specially designed band with 2 built-in side electrodes for recording 1-lead ECG [63]. The sale of Kardia Band was terminated after the Apple Watch Series 4 was released.

Istel HR-2000 (Istel ECG)

Istel HR-2000 (Diagnosis SA) is a CE-certified device that has 4 built-in electrodes corresponding to 5 electrodes of a conventional ECG: the left arm, right arm, left leg, and right leg. The system records a 6-lead real-time ECG and an automatic algorithm recognizes AF. High-quality reports might be analyzed by experts if the result is ambiguous. The 6-lead ECG might serve as an event recorder, thus allowing the identification of other types of arrhythmia, like supraventricular or ventricular tachycardia, premature ventricular or supraventricular contractions, and atrioventricular blocks [64]. No specificity or sensitivity values for AF detection or the correlation status between intervals measured by the device and conventional ECG has been validated in clinical trials.

ECG App for Apple Watch: Series 4 or Subsequent

In 2018, Apple Incorporated introduced the Apple Watch Series 4, the first smartwatch to record a 1-lead ECG, corresponding with lead I from conventional ECG. Apple Watch Series 4 included 2 black crystal electrodes on the back and another electrode that serves as a Digital Crown [65]. An Apple-sponsored multicenter study with 588 patients was performed to determine the Health app's ability to generate an ECG curve corresponding to lead I from a conventional ECG and to use an algorithm classifying heart rhythms as either a sinus rhythm or AF [66]. The results were quite promising: the sensitivity for AF detection was 98.3% and the specificity was 99.6%. Consequently, the app was awarded FDA approval for Apple Watch Series 4 and above [66].

One registered clinical trial in the Cleveland Clinic has compared to the Apple Watch Series 4 and standard telemetry monitoring [67]. Recruitment has finished, but the publication of results is still pending.

Other Technologies

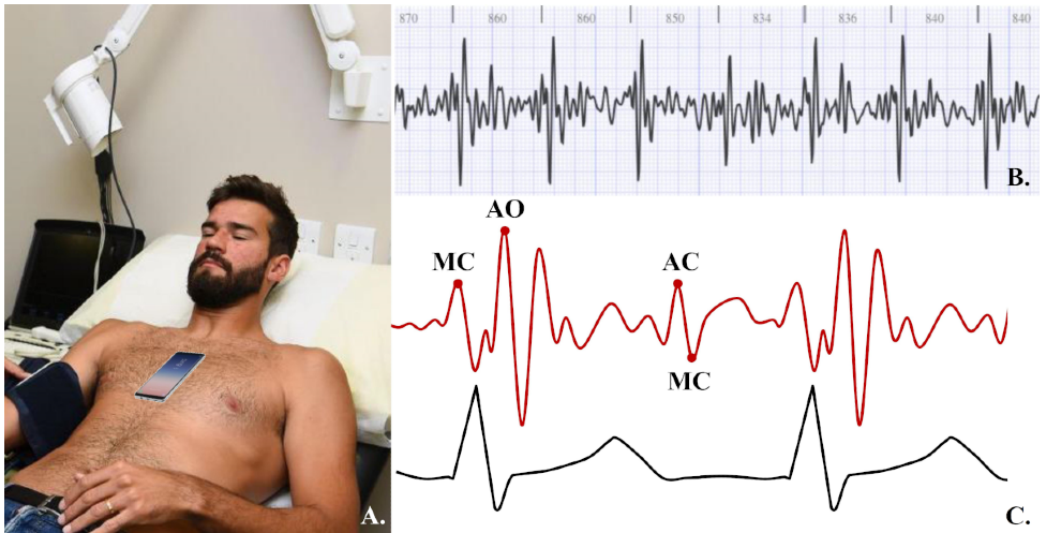
An App Using Phonocardiography: Heart Beat

Heart Beat is a free app that records heart rate using PCG [68]. PCG is a diagnostic technique that records cardiac acoustic phenomena [69] generated by interactions between the blood flow and heart chambers, valves, and great vessels [70]. A microphone must be placed on the chest to correctly measure the heart rate, with the surroundings remaining in absolute silence. The Heart Beat transforms the audio signal into heartbeat frequency [71]. Heart Beat has a few limitations: it has not been used in any clinical trials, absolute silence is needed when recording the signal, and its status for arrhythmia detection still remains unknown. Due to these limitations, the PCG app cannot be recommended for arrhythmia screening.

An App Using Seismocardiography: BeatScanner

BeatScanner is the only app that uses SCG [72]. The app uses a very sensitive built-in accelerometer and gyroscope sensors in the smartphone to acquire microvibrations of the precordial area in reaction to heartbeats, blood flow, and respiration [73,74]. The vibrations can be studied along the superior-inferior axis (head to foot), the sinister-dexter axis (left to right), and the dorsoventral axis (back to front) [75]. The typical signal received by the gyroscope or accelerometer is called a seismocardiogram. The peaks in the seismocardiogram correspond to the opening and closing of the mitral and aortic valve [74,76]. The averaged SCG signal corresponds to ECG (Figure 4) [72,74,76]. According to Salerno and Zanetti [77], SCG might be applied to monitor the function of the left ventricle during ischemia. Paukkunen et al [78] propose that SCG may play a role in detecting atrial flutter. SCG may prove to be useful in arrhythmia detection, as the sensors are built into devices such as smartphones, and the method is noninvasive. Moreover, the sensors are cheap to develop, and the obtained signal is of high quality [75]. Unfortunately, no randomized controlled trials have compared BeatScanner with any of the methods validated for arrhythmia detection.

Figure 4. (A) The method of testing the seismocardiographic signal by smartphone (user in a reclined position or lying down). (B)_A raw seismocardiogram of a normal sinus rhythm presented in the tab "Signal representation" of the BeatScanner app [72]. (C) Correspondence between the averaged seismocardiogram (red curve) and electrocardiogram (black curve); MC describes the main peaks of seismocardiogram signal. Adapted from adapted from Shafiq et al [74]. MC: mitral valve closing; AO: aortic valve opening; AC: aortic valve closing; MO: mitral valve opening.



Current Status and Future of Smartphone Apps in Mobile Health

A number of clinical trials have demonstrated that mobile apps both with and without accessory devices can play a valuable role in arrhythmia screening and that this role may grow in the future. The list of trials given in Table 3 includes those regarding the sensitivity and specificity of the apps and were published in PubMed before June 2020; these studies were identified by a search using the name of the app or name of technology. Some of the apps were evaluated individually (ie, FibriCheck in the DIGITAL AF II at the screening phase) so that the specificity or sensitivity is not available. Others were compared to each other or the gold standard (ie, conventional 12-lead ECG; Table 3).

Table 3. Mobile app in clinical research, including their sensitivity and specificity in detecting atrial fibrillation for individual applications

Study by app examined	Method with which the app was compared	Sensitivity (%)	Specificity (%)
Smartphone apps			
Photo Afib Detector			
Krivoshei et al, 2017 [79]	12-lead ECG[a]	87.5	95.0
McManus et al, 2013 [28]	12-lead ECG	96.2	97.5
Cardiio Rhythm[b]			
Rozen et al, 2018 [32]	12-lead ECG	93.1	90.9
Yan et al, 2018 [33]	12-lead ECG	95.0	96.0
Chan et al, 2016 [31]	AliveCor	92.9	97.7
Preventicus Heartbeats			
Brasier et al (DETECT AF PRO[c]), 2019 [34]	AliveCor	89.9/91.3/91.5[d]	99.1/98.7/99.6[d]
FibriCheck			
Proesmans et al, 2019 [39]	12-lead ECG	96.0	91.1
Verbrugge et al, DIGITAL AF II[e], 2019 [38]	N/A[f]	N/A	N/A
Kardia Mobile[g]			
Selder et al, 2019 [54]	12-lead ECG	92.0	95.0
Koltowski et al, 2019 [61]	12-lead ECG	92.8	100.0
Himmelreich et al, 2019 [56]	12-lead ECG	87.0	97.9
Brasier et al (DETECT AF PRO), 2019 [34]	Preventicus Heartbeats	99.6	97.8
William et al. (iREAD[h]), 2018 [80]	12-lead ECG	96.6	94.1
Lown et al (SAFETY[i]), 2018 [81]	12-lead ECG	97.8	98.8
Chan et al, 2016 [31]	Cardiio Rhythm	71.4	99.4
Lowres et al (SEARCH-AF), 2014 [82]	12-lead ECG	98.5	91.4
Smartwatch apps			
Preventicus Nightwatch			
Dörr et al, (WATCH AF[j]), 2019 [36]	AliveCor	93.7	98.2
Kardia Band			
Wasserlauf et al, 2019 [83]	Reveal LINQ	97.5	N/A
Bumgarner et al, 2018 [84]	12-lead ECG	93.0	84.0
Health for Apple Watch Series 4			
Apple Incorporated, 2018 [85]	12-lead ECG	98.3	99.6

[a]ECG: electrocardiogram.

[b]Beta version not commercially available.

[c]DETECT AF PRO: Enhanced Diagnostics for Early Detection of Atrial Fibrillation–Prospective Validation

[d]Sensitivity and specificity values increased in the course of recording time from 1-3 to 5 minutes.

[e]DIGITAL AF II: Impact of Smartphone-Based Atrial Fibrillation Screening in the General Population for Primary Stroke Prevention.

[f]N/A: not available.

[g]Formerly known as AliveCor.

[h]iREAD: Assessing the Accuracy of an Automated Atrial Fibrillation Detection Algorithm Using Smartphone Technology.

[i]SAFETY: Screening for Atrial Fibrillation Using Economical and Accurate Technology.

[j]WATCH-AF: Smartwatches for Detection of Atrial Fibrillation.

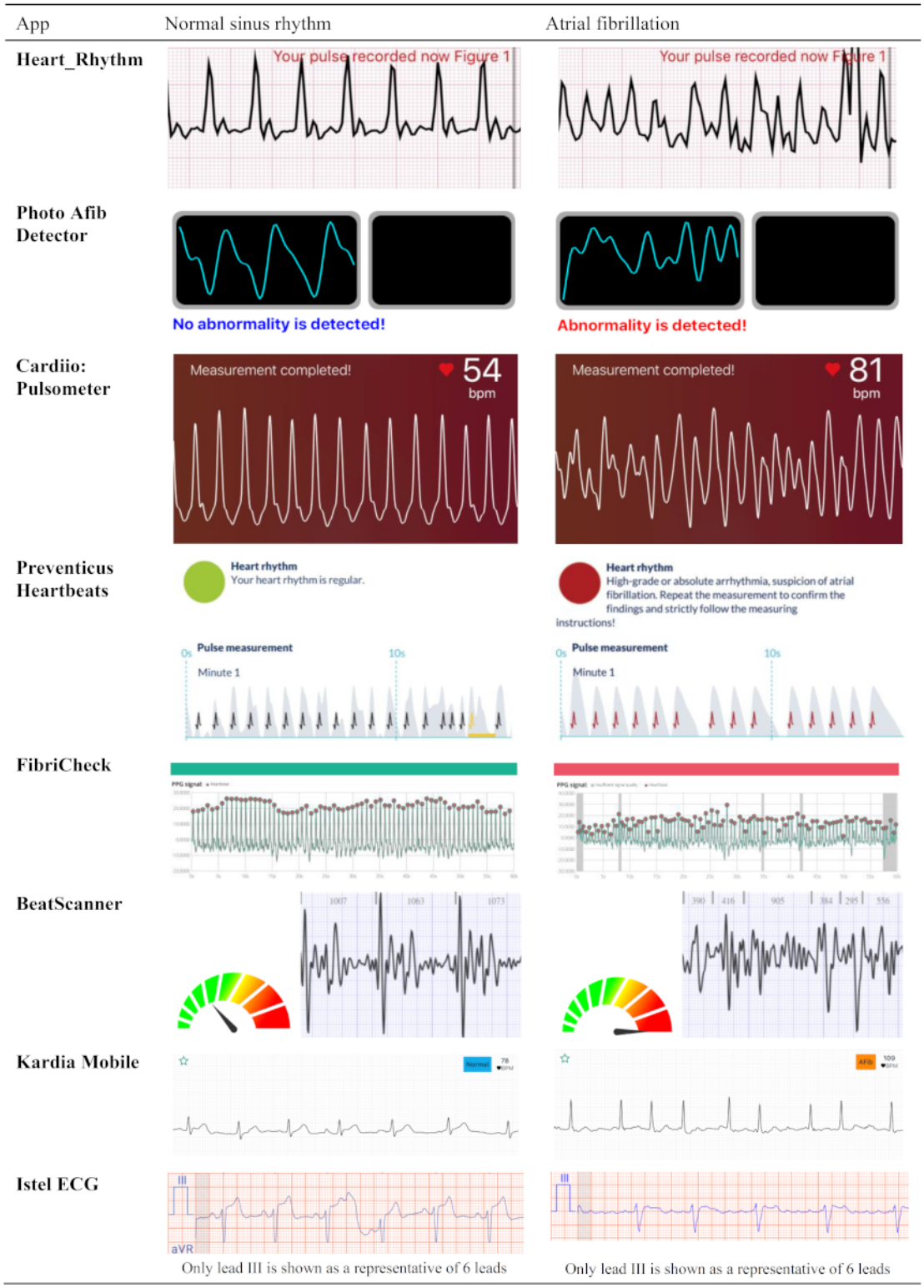
Smartphone or smartwatch apps appear easy to use and are characterized by high accuracy in arrhythmia detection [86]. They may serve as noninvasive event recorders in patients with unexplained palpitations or presyncope [87]. In addition, heart

rhythm monitoring based on AliveCor was well received among the pediatric population compared to conventional telemetry devices [59]. As mHealth components, mobile apps can be effectively used to detect the first episode or early recurrence of atrial arrhythmia in patients with high stroke risk and unknown AF [88] or following ablation or cardioversion [89]. Finally, screening for AF with mobile apps can lower the risk of stroke and reduce the economic burden: its use has a good cost-effectiveness ratio [82,86].

The two leading methods of arrhythmia screening are PPG and iECG, with the former being more accessible. Although PPG still needs further investigation, the results of The Huawei Heart Study [90] and The Apple Heart Study [85], conducted on 187,912 and 419,093 participants respectively, seem promising. The findings indicate that PPG may play a significant role in AF screening by detecting heart rhythm irregularity. Regarding the iECG method, European Heart Rhythm Association findings suggest that clinicians' interpretation of arrhythmia episodes detected by apps does not need to be confirmed with ECG before treatment initiation [91]. Apps based on PCG or SCG face a number of hurdles before implementation due to the substantial interference between chest sounds (in PCG) or oscillations (in SCG) with ambient sound or body tremors, the need for direct access to the chest, the need for complete contact between the phone and the chest wall, and the need for of a compulsory position to perform the measurement. These technical details make PCG or SCG less useful than iECG or PPG in everyday practice. In addition, no PCG or SCG apps have been evaluated thus far in clinical trials.

A combination of technologies, such as PPG with subsets of artificial intelligence, is changing health outcomes worldwide. A summary of normal sinus rhythm and AF reports generated by selected apps is shown in Figure 5.

Figure 5. Summary of normal sinus rhythm and atrial fibrillation reports, generated by selected apps with detection of irregularity, if applicable.



The role of apps supporting AF diagnosis and treatment will doubtlessly grow [90]. Since the first publication regarding the possibility of using PPG in AF detection (McManus et al [28] in 2013), its role has been developed and consolidated. Nowadays, PPG devices are not only used to confirm heart rate or check its regularity, but they can also record real-time iECG and serve as an indication for a specialist to initiate treatment [88]. The apps help detect the first episode of AF, monitor the heart rhythm in paroxysmal AF, monitor the heart rate in permanent AF, and connect the symptom with other arrhythmias or conduction abnormalities [92].

During the 2019 COVID-19 pandemic, when face-to-face consultations were transformed into teleconsultations, the value of smartphone apps and mHealth solutions in remote arrhythmia management was confirmed [93]. With the pandemic gathering pace, mobile apps will undoubtedly become a more fixed part of health infrastructure.

In addition to arrhythmia screening, some apps can be used for other applications. The literature has discussed the potential for detecting real-time myocardial ischemia using single-lead Kardia Mobile [94] or even ST-elevated myocardial infarction of the inferior wall by transforming single-lead Apple Watch Series 4 into a triple-lead smartwatch [95]. Also, the newly introduced Kardia Mobile 6L seems to be a perfect device for diagnosing myocardial ischemia and even myocardial infarction of the inferior wall, owing to its 6-lead ECG feature [62]. In addition, Yasin et al [96] found that an iECG signal could be processed to calculate the serum potassium concentration in patients undergoing hemodialysis.

Conclusions

The rapid development of mHealth solutions may revolutionize approaches to arrhythmia screening. The ECG- and PPG-based apps demonstrate greater availability and efficacy in AF detection than those using PCG or SCG.

ECG apps can be used to detect AF; in addition, the results can also be used to precisely diagnose other types of arrhythmias (narrow or wide QRS complex tachycardia, premature supraventricular or ventricular contractions), conduction abnormalities (atrioventricular blocks, intraventricular blocks of undetermined origin), and pathological intervals (short or long QT) if the ECG trace is interpreted by a specialist. In contrast, PPG apps can be used to detect AF or to diagnose general tachycardia or bradycardia of undetermined etiology or premature contractions of undetermined origin. Therefore, it is recommended that PPG apps be used for monitoring treatment efficacy and that ECG apps be used for determining a diagnosis of AF, as robust traces are essential to starting proper treatments, such as those that included oral anticoagulants. However, due to technical details and lack of evidence, PCG or SCG apps cannot be recommended for setting a diagnosis of AF or for monitoring treatment efficacy.

As new technologies are still being developed, clinical trials of mobile apps in health care are ongoing. The findings obtained so far suggest they are on the right track to improving the efficacy of early detection of AF, thus lowering the risk of stroke and reducing the economic burden placed on public health.

Acknowledgments

None declared.

Conflicts of Interest

None declared.

References

1. Wyndham CR. Atrial fibrillation: the most common arrhythmia. Tex Heart Inst J 2000;27(3):257-267 [FREE Full text] [Medline: 11093410]
2. Fitzmaurice DA, Hobbs FDR, Jowett S, Mant J, Murray ET, Holder R, et al. Screening versus routine practice in detection of atrial fibrillation in patients aged 65 or over: cluster randomised controlled trial. BMJ 2007 Aug 25;335(7616):383 [FREE Full text] [doi: 10.1136/bmj.39280.660567.55] [Medline: 17673732]
3. Zoni-Berisso M, Lercari F, Carazza T, Domenicucci S. Epidemiology of atrial fibrillation: European perspective. Clin Epidemiol 2014;6:213-220 [FREE Full text] [doi: 10.2147/CLEP.S47385] [Medline: 24966695]
4. Wolf PA, Dawber TR, Thomas HE, Kannel WB. Epidemiologic assessment of chronic atrial fibrillation and risk of stroke: the Framingham study. Neurology 1978 Oct;28(10):973-977. [doi: 10.1212/wnl.28.10.973] [Medline: 570666]
5. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, ESC Scientific Document Group. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. Eur Heart J 2021 Feb 01;42(5):373-498. [doi: 10.1093/eurheartj/ehaa612] [Medline: 32860505]
6. Rho RW, Page RL. Asymptomatic atrial fibrillation. Prog Cardiovasc Dis 2005;48(2):79-87. [doi: 10.1016/j.pcad.2005.06.005] [Medline: 16253649]
7. Blackshear JL, Safford RE. Immediate and long-term prevention of thromboembolism in atrial fibrillation. Cardiac Electrophysiol Review 1999;3:88-92. [doi: 10.1023/A:1009949007486]
8. O'Dea S. Number of smartphone users worldwide from 2016 to 2021 (in billions). Statista. URL: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide [accessed 2021-03-15]
9. Vailshery LS. Number of connected wearable devices worldwide from 2016 to 2022. Statista. URL: https://www.statista.com/statistics/487291/global-connected-wearable-devices/ [accessed 2021-03-15]
10. Big data and wearable health monitors: harnessing the benefits and overcoming challenges. Health Informatics. 2019 Sep 17. URL: https://healthinformatics.uic.edu/blog/big-data-and-wearable-health-monitors-harnessing-the-benefits-and-overcoming-challenges/ [accessed 2021-03-15]

11. Shan R, Ding J, Plante TB, Martin SS. Mobile health access and use among individuals with or at risk for cardiovascular disease: 2018 health information national trends survey (hints). J Am Heart Assoc 2019 Dec 17;8(24):e014390 [FREE Full text] [doi: 10.1161/JAHA.119.014390] [Medline: 31818220]
12. Pulse. Saga. URL: https://www.saga.co.uk/contentlibrary/saga/publishing/verticals/health-and-wellbeing/exercise-and-fitness/pulse768x576.jpg [accessed 2019-09-18]
13. Standard systems of heart rhythm monitoring. Mistry Medical Supplies. URL: https://www.mistrymedical.com/graphics/products/large/d4z7q9.jpg [accessed 2019-09-18]
14. kardio.CLINIC. URL: http://kardio.clinic/wp-content/uploads/2016/04/holter-300x300.jpg [accessed 2019-09-18]
15. Rho R, Vossler M, Blancher S, Poole JE. Comparison of 2 ambulatory patch ECG monitors: The benefit of the P-wave and signal clarity. Am Heart J 2018 Sep;203:109-117. [doi: 10.1016/j.ahj.2018.03.022] [Medline: 30072027]
16. Konka A, Sredniawa B. Nowoczesne metody stosowane u pacjentów w celu przesiewowego wykrywania migotania przedsionków. Medycyna po Dyplomie 2019;28:54-56.
17. Implantable loop recorder insertion. Ruth Kam Heart & Arrhthmia Clinic. URL: http://www.arrhythmia.com.sg/wp-content/uploads/2016/07/implantable_loop_recorder_insertion_img_01.jpg [accessed 2019-09-21]
18. MyDiagnostick. URL: https://www.mydiagnostick.com [accessed 2019-09-20]
19. The Heart Check. URL: https://www.theheartcheck.com [accessed 2019-09-20]
20. AfibAlert. Lohman Tech. URL: https://www.lohmantech.com [accessed 2019-09-20]
21. Elgendi M. On the analysis of fingertip photoplethysmogram signals. Curr Cardiol Rev 2012 Feb;8(1):14-25 [FREE Full text] [doi: 10.2174/157340312801215782] [Medline: 22845812]
22. Raj R, Jothi S. Estimation of heart rate from photoplethysmographic signal using SVR method. The International Journal of Science & Technology, Volume, Issue 2014;2(2):50-52.
23. Tamura T, Maeda Y, Sekine M, Yoshida M. Wearable Photoplethysmographic Sensors—Past and Present. Electronics 2014 Apr 23;3(2):282-302. [doi: 10.3390/electronics3020282]
24. Measuring heart rate with mobile photoplethysmography. Squarespace. URL: https://images.squarespace-cdn.com/content/5397da1ae4b0c0c130832759/1460100389675-GROLYC8RNTTA536ZDIOV/?format=1000w&content-type=image%2Fpng [accessed 2019-09-22]
25. Your heart's health at your fingertips. FibriCheck. URL: https://www.fibricheck.com [accessed 2019-09-22]
26. Heart_Rhythm. Apple App Store. URL: https://apps.apple.com/us/app/heart-rhythm/id772332884 [accessed 2019-09-22]
27. Photo AFib Detector (Lite). Apple App Store. URL: https://apps.apple.com/us/app/photo-afib-detector-free/id1002024668 [accessed 2019-09-22]
28. McManus DD, Lee J, Maitas O, Esa N, Pidikiti R, Carlucci A, et al. A novel application for the detection of an irregular pulse using an iPhone 4S in patients with atrial fibrillation. Heart Rhythm 2013 Mar;10(3):315-319 [FREE Full text] [doi: 10.1016/j.hrthm.2012.12.001] [Medline: 23220686]
29. Yan BP, Chan CK, Li CK, To OT, Lai WH, Tse G, et al. Resting and postexercise heart rate detection from fingertip and facial photoplethysmography using a smartphone camera: a validation study. JMIR Mhealth Uhealth 2017 Mar 13;5(3):e33 [FREE Full text] [doi: 10.2196/mhealth.7275] [Medline: 28288955]
30. Bending Spoons. URL: https://bendingspoons.com/support.html?app=542891434#5d8091b2fa90cfc74f2c09dc [accessed 2019-09-22]
31. Chan P, Wong C, Poh YC, Pun L, Leung WW, Wong Y, et al. Diagnostic performance of a smartphone-based photoplethysmographic application for atrial fibrillation screening in a primary care setting. J Am Heart Assoc 2016 Jul 21;5(7):e003428 [FREE Full text] [doi: 10.1161/JAHA.116.003428] [Medline: 27444506]
32. Rozen G, Vaid J, Hosseini SM, Kaadan MI, Rafael A, Roka A, et al. Diagnostic accuracy of a novel mobile phone application for the detection and monitoring of atrial fibrillation. Am J Cardiol 2018 May 15;121(10):1187-1191. [doi: 10.1016/j.amjcard.2018.01.035] [Medline: 29525063]
33. Yan BP, Lai WHS, Chan CKY, Chan SC, Chan L, Lam K, et al. Contact-free screening of atrial fibrillation by a smartphone using facial pulsatile photoplethysmographic signals. J Am Heart Assoc 2018 Apr 05;7(8):e [FREE Full text] [doi: 10.1161/JAHA.118.008585] [Medline: 29622592]
34. Brasier N, Raichle CJ, Dörr M, Becke A, Nohturfft V, Weber S, et al. Detection of atrial fibrillation with a smartphone camera: first prospective, international, two-centre, clinical validation study (DETECT AF PRO). Europace 2019 Jan 01;21(1):41-47 [FREE Full text] [doi: 10.1093/europace/euy176] [Medline: 30085018]
35. Preventicus. URL: https://en.preventicus.com/en/news-ehra-late-breaking-science-sessions-atrial-fibrillation-af-can-captured-smartphone-cameras-smartwatches-95-accuracy/ [accessed 2019-09-25]
36. Dörr M, Nohturfft V, Brasier N, Bosshard E, Djurdjevic A, Gross S, et al. The WATCH AF Trial: SmartWATCHes for Detection of Atrial Fibrillation. JACC Clin Electrophysiol 2019 Feb;5(2):199-208. [doi: 10.1016/j.jacep.2018.10.006] [Medline: 30784691]
37. Vandenberk T, Stans J, Van SG, Pelckmans C, Smeets CJ, Lanssens D, et al. Clinical validation of heart rate apps: mixed-methods evaluation study. JMIR Mhealth Uhealth 2017 Aug 25;5(8):e129 [FREE Full text] [doi: 10.2196/mhealth.7254] [Medline: 28842392]

38. Verbrugge F, Proesmans T, Vijgen J, Mullens W, Rivero-Ayerza M, Van Herendael H, et al. Atrial fibrillation screening with photo-plethysmography through a smartphone camera. Europace 2019 Aug 01;21(8):1167-1175. [doi: 10.1093/europace/euz119] [Medline: 31056678]
39. Proesmans T, Mortelmans C, Van Haelst R, Verbrugge F, Vandervoort P, Vaes B. Mobile phone–based use of the photoplethysmography technique to detect atrial fibrillation in primary care: diagnostic accuracy study of the fibricheck app. JMIR Mhealth Uhealth 2019 Mar 27;7(3):e12284. [doi: 10.2196/12284] [Medline: 30916656]
40. FibriCheck. Apple App Store. URL: https://apps.apple.com/be/app/fibricheck/id1102327473 [accessed 2019-09-22]
41. KardiaMobile phone clip. AliveCor. URL: https://store.alivecor.com/products/kardiamobile-phoneclip [accessed 2019-10-02]
42. KardiaMobile 6L. level. URL: https://leveldesignsf.com/alivecor [accessed 2019-10-02]
43. Ergodesk. URL: https://www.ergodesk.cz/en/products/biofeedback/alivecor-kardia-ecg-band-apple-watch [accessed 2019-10-02]
44. Istel. URL: http://istel.com.pl/rejestrator-ekg/?lang=en [accessed 2019-10-02]
45. Mobile ECG Event Recorder Heart Monitor ECG Check APP for family use. Borsam. URL: http://ecg-borsam-com.sell.everychina.com/p-105202847-mobile-ecg-event-recorder-heart-monitor-ecg-check-app-for-family-use.html [accessed 2019-10-02]
46. PROAlleine. iF Design. URL: https://uploads.ifdesign.de/award_img_247/oex_large/107521_02_rendering_PROalleine.jpg [accessed 2020-09-20]
47. Agatsa. URL: https://www.agatsa.com/wp-content/uploads/2018/07/device.png [accessed 2020-09-20]
48. CardioComm. Mobile Health News. URL: https://www.mobihealthnews.com/sites/default/files/cardiocomm_0.png [accessed 2020-09-20]
49. Cardiology 2.0. URL: https://magazine.cardiology2.com/cardio2/feb2020/coalalife/397672/Coala_Pro.960_0_1.jpg [accessed 2020-09-20]
50. Medical Expo. URL: https://img.medicalexpo.com/images_me/photo-g/84877-8973365.jpg [accessed 2020-09-20]
51. ECG app for Apple Watch will reportedly launch with watchOS 5.1.2. iDownloadBlog. URL: https://www.idownloadblog.com/2018/11/28/watchos-5-1-2-ecg-app/ [accessed 2019-10-02]
52. AliveCor. KardiaMobile. URL: https://www.alivecor.com/kardiamobile/ [accessed 2019-10-02]
53. Garabelli P, Albert D. Innovation Abstract # 12-IA-9662-HRS. 2012 Presented at: Heart Rhythm Society's 33rd Annual Scientific Sessions; 2012 May 9-12; Boston, Massachusetts. [doi: 10.1016/s1547-5271(15)00403-8]
54. Selder JL, Breukel L, Blok S, van Rossum AC, Tulevski II, Allaart CP. A mobile one-lead ECG device incorporated in a symptom-driven remote arrhythmia monitoring program. The first 5,982 Hartwacht ECGs. Neth Heart J 2019 Jan;27(1):38-45 [FREE Full text] [doi: 10.1007/s12471-018-1203-4] [Medline: 30523617]
55. Krajsman M, Polinski J, Pawlik K, Tataj E, Parol G, Cacko A. Utility of mobile single-lead ECG device in hospital emergency department. Folia Cardiologica 2018;5:402-406. [doi: 10.5603/fc.2018.0110]
56. Himmelreich JC, Karregat EP, Lucassen WA, van Weert HC, de Groot JR, Handoko ML, et al. Diagnostic accuracy of a smartphone-operated, single-lead electrocardiography device for detection of rhythm and conduction abnormalities in primary care. Ann Fam Med 2019 Sep;17(5):403-411 [FREE Full text] [doi: 10.1370/afm.2438] [Medline: 31501201]
57. Garabelli P, Stavrakis S, Albert M, Koomson E, Parwani P, Chohan J, et al. Comparison of QT interval readings in normal sinus rhythm between a smartphone heart monitor and a 12-lead ECQ for healthy volunteers and inpatients receiving sotalol or dofetilide. J Cardiovasc Electrophysiol 2016 Jul;27(7):827-832. [doi: 10.1111/jce.12976] [Medline: 27027653]
58. Abudan AA, Isath A, Ryan JD, Henrich MJ, Dugan JL, Attia ZI, et al. Safety and compatibility of smart device heart rhythm monitoring in patients with cardiovascular implantable electronic devices. J Cardiovasc Electrophysiol 2019 Sep;30(9):1602-1609. [doi: 10.1111/jce.14024] [Medline: 31190453]
59. Nguyen HH, Van Hare GF, Rudokas M, Bowman T, Silva JNA. SPEAR Trial: Smartphone Pediatric ElectrocARdiogram Trial. PLoS One 2015;10(8):e0136256 [FREE Full text] [doi: 10.1371/journal.pone.0136256] [Medline: 26295569]
60. Gropler MRF, Dalal AS, Van Hare GF, Silva JNA. Can smartphone wireless ECGs be used to accurately assess ECG intervals in pediatrics? A comparison of mobile health monitoring to standard 12-lead ECG. PLoS One 2018;13(9):e0204403 [FREE Full text] [doi: 10.1371/journal.pone.0204403] [Medline: 30260996]
61. Koltowski L, Balsam P, Glowczynska R, Rokicki JK, Peller M, Maksym J, et al. Kardia Mobile applicability in clinical practice: A comparison of Kardia Mobile and standard 12-lead electrocardiogram records in 100 consecutive patients of a tertiary cardiovascular care center. Cardiol J 2021;28(4):543-548 [FREE Full text] [doi: 10.5603/CJ.a2019.0001] [Medline: 30644079]
62. KardiaMobile6L. KardiaMobile. URL: https://www.alivecor.com/kardiamobile6l/ [accessed 2019-10-02]
63. Kardia Band by AliveCor instructions for use. KardiaMobile. URL: https://www.alivecor.com/previous-labeling/kb_ifu/15LB1.3.pdf [accessed 2019-10-02]
64. Istel. URL: http://istel.com.pl/ [accessed 2019-10-05]
65. ECG app for Apple Watch will reportedly launch with watchOS 5.1.2. iDownload Blog. URL: https://www.idownloadblog.com/2018/11/28/watchos-5-1-2-ecg-app/ [accessed 2019-10-05]
66. Apple App store. URL: https://www.apple.com/healthcare/site/docs/Apple_Watch_Arrhythmia_Detection.pdf [accessed 2019-10-05]

67. Gillinov M. Accuracy of the Apple Watch Series 4 for detection of heart rhythm: a pilot study. ClinicalTrials.gov. URL: https://clinicaltrials.gov/ct2/show/NCT03798613?cond=apple+watch+4&draw=2&rank=1 [accessed 2019-10-05]
68. Google Play. URL: https://play.google.com/store/apps/details?id=com.hahoteknik.HeartBeatMaster&hl=en_US [accessed 2019-09-22]
69. Noponen A, Lukkarinen S, Sepponen R. Section I: current methods and their applications for cardiovascular multimodal imaging. Chapter 2: phonocardiography. In: Pahlm O, Wagner GS, editors. Multimodal Cardiovascular Imaging: Principles and Clinical Applications. New York: The McGraw-Hill Companies; 2011.
70. Tang H, Zhang J, Sun J, Qiu T, Park Y. Phonocardiogram signal compression using sound repetition and vector quantization. Comput Biol Med 2016 Apr 01;71:24-34 [FREE Full text] [doi: 10.1016/j.compbiomed.2016.01.017] [Medline: 26871603]
71. Heart Beat Rate Monitor App. YouTube. URL: https://www.youtube.com/watch?v=9GFThK8Iu0k [accessed 2019-09-22]
72. Features. BeatScanner. URL: https://www.beatscanner.com/#features [accessed 2019-09-28]
73. Jafari Tadi M, Lehtonen E, Saraste A, Tuominen J, Koskinen J, Teräs M, et al. Gyrocardiography: a new non-invasive monitoring method for the assessment of cardiac mechanics and the estimation of hemodynamic variables. Sci Rep 2017 Jul 28;7(1):6823 [FREE Full text] [doi: 10.1038/s41598-017-07248-y] [Medline: 28754888]
74. Shafiq G, Tatinati S, Ang WT, Veluvolu KC. Automatic identification of systolic time intervals in seismocardiogram. Sci Rep 2016 Nov 22;6:37524 [FREE Full text] [doi: 10.1038/srep37524] [Medline: 27874050]
75. Wahlstrom J, Skog I, Handel P, Khosrow-Khavar F, Tavakolian K, Stein PK, et al. A hidden Markov Model for seismocardiography. IEEE Trans Biomed Eng 2017 Oct;64(10):2361-2372. [doi: 10.1109/TBME.2017.2648741] [Medline: 28092512]
76. Laurin A, Imperiale S, Moireau P, Blaber A, Chapelle D. A 3D model of the thorax for seismocardiography. Computing in Cardiology Conference 2015:465-468. [doi: 10.1109/cic.2015.7408687]
77. Salerno DM, Zanetti J. Seismocardiography for monitoring changes in left ventricular function during ischemia. Chest 1991 Oct;100(4):991-993. [doi: 10.1378/chest.100.4.991] [Medline: 1914618]
78. Paukkunen M, Parkkila P, Hurnanen T, Pänkäälä M, Koivisto T, Nieminen T, et al. Beat-by-beat quantification of cardiac cycle events detected from three-dimensional precordial acceleration signals. IEEE J Biomed Health Inform 2016 Mar;20(2):435-439. [doi: 10.1109/JBHI.2015.2391437] [Medline: 25594987]
79. Krivoshei L, Weber S, Burkard T, Maseli A, Brasier N, Kühne M, et al. Smart detection of atrial fibrillation†. Europace 2017 May 01;19(5):753-757 [FREE Full text] [doi: 10.1093/europace/euw125] [Medline: 27371660]
80. William AD, Kanbour M, Callahan T, Bhargava M, Varma N, Rickard J, et al. Assessing the accuracy of an automated atrial fibrillation detection algorithm using smartphone technology: The iREAD Study. Heart Rhythm 2018 Oct;15(10):1561-1565. [doi: 10.1016/j.hrthm.2018.06.037] [Medline: 30143448]
81. Lown M, Yue AM, Shah BN, Corbett SJ, Lewith G, Stuart B, et al. Screening for Atrial Fibrillation Using Economical and Accurate Technology (From the SAFETY Study). Am J Cardiol 2018 Oct 15;122(8):1339-1344. [doi: 10.1016/j.amjcard.2018.07.003] [Medline: 30131106]
82. Lowres N, Neubeck L, Salkeld G, Krass I, McLachlan AJ, Redfern J, et al. Feasibility and cost-effectiveness of stroke prevention through community screening for atrial fibrillation using iPhone ECG in pharmacies. The SEARCH-AF study. Thromb Haemost 2014 Jun;111(6):1167-1176. [doi: 10.1160/TH14-03-0231] [Medline: 24687081]
83. Wasserlauf J, You C, Patel R, Valys A, Albert D, Passman R. Smartwatch performance for the detection and quantification of atrial fibrillation. Circ Arrhythm Electrophysiol 2019 Jun;12(6):e006834. [doi: 10.1161/CIRCEP.118.006834] [Medline: 31113234]
84. Bumgarner JM, Lambert CT, Hussein AA, Cantillon DJ, Baranowski B, Wolski K, et al. Smartwatch algorithm for automated detection of atrial fibrillation. J Am Coll Cardiol 2018 May 29;71(21):2381-2388 [FREE Full text] [doi: 10.1016/j.jacc.2018.03.003] [Medline: 29535065]
85. Turakhia MP, Desai M, Hedlin H, Rajmane A, Talati N, Ferris T, et al. Rationale and design of a large-scale, app-based study to identify cardiac arrhythmias using a smartwatch: The Apple Heart Study. Am Heart J 2019 Jan;207:66-75 [FREE Full text] [doi: 10.1016/j.ahj.2018.09.002] [Medline: 30392584]
86. Giebel GD, Gissel C. Accuracy of mHealth devices for atrial fibrillation screening: systematic review. JMIR Mhealth Uhealth 2019 Jun 16;7(6):e13641 [FREE Full text] [doi: 10.2196/13641] [Medline: 31199337]
87. Reed MJ, Grubb NR, Lang CC, O'Brien R, Simpson K, Padarenga M, et al. Multi-centre randomised controlled trial of a smart phone-based event recorder alongside standard care versus standard care for patients presenting to the Emergency Department with palpitations and pre-syncope - the IPED (Investigation of Palpitations in the ED) study: study protocol for a randomised controlled trial. Trials 2018 Dec 29;19(1):711 [FREE Full text] [doi: 10.1186/s13063-018-3098-1] [Medline: 30594256]
88. Halcox JP, Wareham K, Cardew A, Gilmore M, Barry JP, Phillips C, et al. Assessment of remote heart rhythm sampling using the AliveCor Heart Monitor to screen for atrial fibrillation: The REHEARSE-AF study. Circulation 2017 Nov 07;136(19):1784-1794. [doi: 10.1161/CIRCULATIONAHA.117.030583] [Medline: 28851729]
89. Goldenthal IL, Sciacca RR, Riga T, Bakken S, Baumeister M, Biviano AB, et al. Recurrent atrial fibrillation/flutter detection after ablation or cardioversion using the AliveCor KardiaMobile device: iHEART results. J Cardiovasc Electrophysiol 2019 Nov;30(11):2220-2228 [FREE Full text] [doi: 10.1111/jce.14160] [Medline: 31507001]

90. Guo Y, Wang H, Zhang H, Liu T, Liang Z, Xia Y, MAFA II Investigators. Mobile photoplethysmographic technology to detect atrial fibrillation. J Am Coll Cardiol 2019 Nov 12;74(19):2365-2375 [FREE Full text] [doi: 10.1016/j.jacc.2019.08.019] [Medline: 31487545]
91. Svennberg E, Camm A. Mobile Health as diagnostic or screening tool for arrhythmias: solution or a threat? EHRA webinar. ESC 365. URL: https://esc365.escardio.org/Congress/EHRA-Webinar/EHRA-Webinars-2019/221278-mobile-health-as-diagnostic-or-screening-tool-for-arrhythmias-solution-or-a-threat [accessed 2019-06-25]
92. Turakhia MP, Kaiser DW. Transforming the care of atrial fibrillation with mobile health. J Interv Card Electrophysiol 2016 Oct;47(1):45-50 [FREE Full text] [doi: 10.1007/s10840-016-0136-3] [Medline: 27306552]
93. Pluymaekers N, Hermans ANL, van der Velden RMJ, den Uijl DW, Vorstermans B, Buskes S, et al. On-demand app-based rate and rhythm monitoring to manage atrial fibrillation through teleconsultations during COVID-19. Int J Cardiol Heart Vasc 2020 Jun;28:100533 [FREE Full text] [doi: 10.1016/j.ijcha.2020.100533] [Medline: 32391412]
94. Joseph J, Redwood S. Mobile digital health devices and the diagnosis in real-time of myocardial ischaemia. QJM 2017 May 01;110(5):309-310 [FREE Full text] [doi: 10.1093/qjmed/hcx046] [Medline: 28339806]
95. Avila CO. Novel use of Apple Watch 4 to obtain 3-lead electrocardiogram and detect cardiac ischemia. Perm J 2019;23:99-102 [FREE Full text] [doi: 10.7812/TPP/19-025] [Medline: 31314734]
96. Yasin OZ, Attia Z, Dillon JJ, DeSimone CV, Sapir Y, Dugan J, et al. Noninvasive blood potassium measurement using signal-processed, single-lead ecg acquired from a handheld smartphone. J Electrocardiol 2017;50(5):620-625 [FREE Full text] [doi: 10.1016/j.jelectrocard.2017.06.008] [Medline: 28641860]

Abbreviations

AF: atrial fibrillation
CE: Conformité Européenne
DETECT AF PRO: Enhanced Diagnostics for Early Detection of Atrial Fibrillation Prospective Validation
DIGITAL AF II: Impact of Smartphone-Based Atrial Fibrillation Screening in the General Population for Primary Stroke Prevention
ECG: electrocardiogram
ESC: European Society of Cardiology
FDA: Food and Drug Administration
iECG: internet-enabled mobile electrocardiogram
mHealth: mobile health
NPV: negative predictive value
PCG: phonocardiography
PPG: photoplethysmography
PPV: positive predictive value
QTc: corrected QT
SCG: seismocardiography
WATCH AF: Smartwatches for Detection of Atrial Fibrillation

Edited by L Buis; submitted 11.08.20; peer-reviewed by M Doerr, C Smeets, E Ding, S Li, J Stans; comments to author 02.09.20; revised version received 04.01.21; accepted 28.07.21; published 27.08.21.

Please cite as:
Książczyk M, Dębska-Kozłowska A, Warchoł I, Lubiński A
Enhancing Healthcare Access–Smartphone Apps in Arrhythmia Screening: Viewpoint
JMIR Mhealth Uhealth 2021;9(8):e23425
URL: https://mhealth.jmir.org/2021/8/e23425
doi:10.2196/23425
PMID:34448723

Original Paper

Effect of Physician-Pharmacist Participation in the Management of Ambulatory Cancer Pain Through a Digital Health Platform: Randomized Controlled Trial

Lu Zhang[1,2*], MD; Howard L McLeod[3,4,5,6*], Prof Dr; Ke-Ke Liu[1,2], MD; Wen-Hui Liu[7], Prof Dr; Hang-Xing Huang[1,2], MD; Ya-Min Huang[1,2], MD; Shu-Sen Sun[1,2,8], Prof Dr; Xiao-Ping Chen[2,5,6], Prof Dr; Yao Chen[5], Prof Dr; Fang-Zhou Liu[9], Prof Dr; Jian Xiao[1,2], Prof Dr

[1]Department of Pharmacy, Xiangya Hospital, Central South University, Changsha, China

[2]Institute for Rational and Safe Medication Practices, National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China

[3]Geriatric Oncology Consortium, Tampa, FL, United States

[4]Taneja College of Pharmacy, University of South Florida, Tampa, FL, United States

[5]Department of Clinical Pharmacology, Xiangya Hospital, Central South University, Changsha, China

[6]Hunan Key Laboratory of Pharmacogenetics, Institute of Clinical Pharmacology, Central South University, Changsha, China

[7]Department of Pharmacy, The Second Xiangya Hospital, Central South University, Changsha, China

[8]College of Pharmacy and Health Sciences, Western New England University, Boston, MA, United States

[9]College of Information Science and Engineering, Hunan Normal University, Changsha, China

[*]these authors contributed equally

Corresponding Author:
Jian Xiao, Prof Dr
Department of Pharmacy
Xiangya Hospital
Central South University
Room 404
No. 87 Xiangya Road, Kaifu District
Changsha, 410008
China
Phone: 86 13973168895
Fax: 86 073184327460
Email: admanoas@163.com

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2021/9/e33223

Abstract

Background: Self-management of ambulatory cancer pain is full of challenges. Motivated by the need for better pain management, we developed a WeChat-supported platform, Medication Housekeeper (MediHK), to enhance communication, optimize outcomes, and promote self-management in the home setting.

Objective: We conducted a randomized controlled trial to assess whether the joint physician-pharmacist team through MediHK would provide better self-management of ambulatory patients with cancer pain.

Methods: Patients were randomly assigned to either an intervention group or control group. During the 4-week study period, the pharmacist would send 24-hour pain diaries daily, adverse drug reaction (ADR) forms every 3 days, and the Brief Pain Inventory form every 15 days to patients in the intervention group via MediHK. If a patient needed a change in drug/dosage or treatment of an ADR after the comprehensive review, the pharmacist would propose pharmacological interventions to the attending physician, who was then responsible for prescribing or adjusting pain medications. If no adjustments were needed, the pharmacist provided appropriate targeted education based on knowledge deficits. Patients in the control group received conventional care

and did not receive reminders to fill out the forms. However, if the control group patients filled out a form via MediHK, the pain management team would review and respond in the same way as for the intervention group. The primary outcomes included pain intensity and pain interference in daily life. Secondary outcomes included patient-reported outcome measures, medication adherence, ADRs, and rehospitalization rates.

Results: A total of 100 patients were included, with 51 (51%) in the intervention group and 49 (49%) in the control group. The worst pain scores, least pain scores, and average pain scores in the intervention group and the control group were statistically different, with median values of 4 (IQR 3-7) vs 7 (IQR 6-8; *P*=.001), 1 (IQR 0-2) vs 2 (IQR 1-3; *P*=.02), and 2 (IQR 2-4) vs 4 (IQR 3-5; *P*=.001), respectively, at the end of the study. The pain interference on patients' general activity, mood, relationships with others, and interests was reduced, but the difference was not statistically significant compared with the control group (*Ps*=.10-.76). The medication adherence rate increased from 43% to 63% in the intervention group, compared with an increase of 33% to 51% in the control group (*P*<.001). The overall number of ADRs increased at 4 weeks, and more ADRs were monitored in the intervention group (*P*=.003). Rehospitalization rates were similar between the 2 groups.

Conclusions: The joint physician-pharmacist team operating through MediHK improved pain management. This study supports the feasibility of integrating the internet into the self-management of cancer pain.

Trial Registration: Chinese Clinical Trial Registry ChiCTR1900023075; https://www.chictr.org.cn/showproj.aspx?proj=36901

(JMIR Mhealth Uhealth 2021;9(8):e24555) doi:10.2196/24555

KEYWORDS

cancer pain; self-management; ambulatory setting; digital health; physician-pharmacist

Introduction

Pain is a common and challenging issue for cancer patients, affecting most at some stage of their disease [1]. A meta-analysis indicated that pain prevalence was 33% in patients after curative treatment, 64% in patients with advanced disease, and 59% in patients on anticancer therapy; approximately 35% of patients reported their pain as moderate to severe [2]. Inadequate pain management continues, with approximately one-third of cancer pain patients undertreated [3]. According to recent surveys, cancer pain management in China remains far from the ideal goal [4]. The barriers are multifactorial, including knowledge deficits, inadequate pain assessment, and misconceptions of pain from both patients and professionals [5]. Managing ambulatory patients is especially tough because of the complex environment, limited communication with health care providers, and difficulty managing their pain-medication regimens [6,7].

Both the World Health Organization and the European Society for Medical Oncology recommended that cancer pain patients should be active in their self-management of their pain. Patient-reported outcomes are increasingly used in routine ambulatory cancer care to guide clinical decisions, enhance communication, and improve symptom management [8]. Electronic patient-reported outcomes, supported by computer-adaptive testing technology, have shown potential in the era of big data [9]. Smartphones and apps such as WeChat (the largest social networking app in China), provide additional value to obtain knowledge and information, as well as making it possible for patients and health care providers to communicate electronically. Most patients are willing to self-report their symptoms via digital health apps. Several studies have reported on applications based on the eHealth model for the self-management of cancer pain [10,11].

This study established a physician-pharmacist collaboration team that participated in the self-management of ambulatory patients with cancer pain through a WeChat-supported platform: Medication Housekeeper (MediHK). We aimed to assess whether the joint physician-pharmacist team operating through MediHK would provide better self-management of ambulatory patients with cancer pain and optimize therapeutic outcomes.

Methods

WeChat-Supported Platform: MediHK Design

Patients were managed by MediHK, a WeChat-supported platform designed by the research team. An engineer from Hunan Normal University's College of Information Science and Engineering provided technology support for building MediHK. MediHK has been patented by the National Intellectual Property Administration, People's Republic of China (ZL 2015 1 0648320.2). MediHK contains 2 opening screens: (1) the patient interface (Multimedia Appendix 1) and (2) the medical interface (Multimedia Appendix 2). The former is for patients, and the latter is for the members of the pain management team, which consisted of physicians and pharmacists. The medical interface was designed to manage pain-related problems and provide consultation to patients in a timely fashion. Both interfaces included 3 modules: a user login module for inserting basic user demographics into MediHK; an e-consultation module for communicating between patients and medical providers; and an introductory module for MediHK education, which offers a quick response code for new users (Multimedia Appendix 3).

We conducted 3 rounds of consensus-building using Delphi methods [12] to determine patient-reported outcome measures (PROMs) that could be integrated into MediHK. The pharmacists sent messages to patients, as shown on the far-right side of Multimedia Appendix 2, and the patients would receive a reminder, as shown on the far-right side of Multimedia Appendix 1. Patients could consult on any questions about pain, and the pain management team would receive real-time WeChat messages and respond as soon as possible. The acceptable response time was generally within 2 hours. When patients needed a change in drug/dosage or treatment of an adverse drug

reaction (ADR), pharmacists first reviewed the patients' historical records through MediHK and, then, made recommendations and reminded the physician. In China, pharmacists have no right to prescribe. If the physician had conflicting opinions, an agreement would be reached through offline contact; then, the physician could adjust drug-therapy regimens. All treatment recommendations from the pharmacists and physicians were based on the guidelines of the European Society for Medical Oncology Standard diagnosis and treatment of cancer pain, the National Comprehensive Cancer Network, and the European Association for Palliative Care.

When the patient's expression was not clear, there was no guarantee that the same physician or pharmacist would communicate back. However, since the previously submitted forms and consultation questions were available through MediHK, the new physician or pharmacist would review the patient's history in all aspects. Patient information was protected by encryption. Note that all screenshots of the app include translations that have been added for clarity for this paper.

PROMs and Forms That Integrated Into the MediHK

The Brief Pain Inventory

This study used the Brief Pain Inventory (BPI; Chinese version) to assess cancer pain, and it has been widely used for its good construct and concurrent validity [13]. It provides 2 main scores, which are "pain intensity" and "pain interference in daily life." Pain intensity is based on the Numerical Rating Scale (NRS) and includes a 4-item sensory dimension: worst pain, least pain, average pain, and present pain. Each item is rated from 0 ("no pain") to 10 ("very severe pain"). The "pain interference on daily life" score is a 7-item reactive dimension that includes general activity, mood, walking ability, daily work, relationships, sleep, and enjoyment of life; each item is rated from 0 ("no interference on daily life") to 10 ("complete interference").

Medication History and Adherence

We designed a list to record the medication history of ambulatory patients with cancer pain (Multimedia Appendix 4). The Morisky Medication Adherence Measure [14] was used to assess adherence to analgesics because of its excellent reliability and validity in the Chinese cancer pain population [15]. The Morisky Medication Adherence Measure focuses on the following medication-taking behaviors (Multimedia Appendix 5): forgetfulness, carelessness, and cessation of the drug regimen when feeling better or worse. The answers of "yes" or "no" for each item scored 0 and 1, respectively. The scores were divided into 3 categories: complete adherence (4), incomplete adherence (1-3), and nonadherence (0).

ADR Form

Many patients treated with opioids may experience adverse events. To comprehensively capture ADRs of patients outside of the hospital, we designed a form (Multimedia Appendix 6) with World Health Organization terminology for ADRs and classified ADRs into several symptoms.

Pain Diary for 24-Hour Pain

We designed a pain diary to capture patients' daily pain in the home setting over time. The pain diary included 5 modules (Multimedia Appendix 7): (1) a pain score record, which combined the NRS, Face Pain Scale, and Verbal Rating Scale to assess pain accurately; (2) a form that included the time of administration as well as medication name and dosage for patients taking medication within 24 hours; (3) a module that recorded the specific duration, pain score, treatment status, and new pain location when the NRS was >4; (4) a module that recorded detailed pain information; and (5) a final module that gave suggestions to physicians or pharmacists based on patient feedback.

Study Design and Participants

Study Overview

This was a 2-arm, randomized controlled clinical trial, and the study has been registered at Chictr.org ChiCTR1900023075; https://www.chictr.org.cn/showproj.aspx?proj=36901. Ambulatory patients with cancer pain in a tertiary hospital were included and assigned to either a control group (ie, joined in the MediHK only, no physician-pharmacist active intervention) or an intervention group (ie, MediHK platform plus physician-pharmacist intervention), with an allocation ratio of 1:1 using a random number table. Our pre-experiment included 72 patients who met the criteria for inclusion. The preliminary results showed that the average NRS of patients in the control group was 5.85 (SD 2.442). The average score of patients in the intervention group was expected to be <4. Assuming a type I error of 5%, a type II error of 20%, and considering the design of similar sample content, the sample size required for each group was calculated to be 37 patients. Allowing for 20% attrition, 100 patients (50 participants per group) were planned to be enrolled. The study was conducted in accordance with the Declaration of Helsinki and was approved by the ethics committee of Xiangya Hospital of Central South University (approval number 2017121139). All participants signed an informed consent form before participation.

Care of Patients in the Intervention Group

The pharmacist would send daily 24-hour pain diaries, ADR forms every 3 days, and the BPI form every 15 days via MediHK. The pharmacist would first review patient demographic information, check the form regarding pain intensity and interference in daily life, conduct medication therapy reviews, and review ADRs and medication adherence. If the patient needed a change in drug/dosage or treatment of an ADR after the comprehensive review, the pharmacist would propose pharmacological interventions to the attending physician. The physician was responsible for prescribing or adjusting pain medications. If no changes were needed, the pharmacist provided appropriate targeted education based on patient knowledge deficits.

Care of Patients in the Control Group

Patients in the control group received conventional care. Before the patient was discharged from the hospital, the pharmacist conducted detailed medication education. However, the control group patients did not receive a reminder to fill out the forms.

If they filled out the form via MediHK, the pain management team would also review and respond the same way as for the intervention group.

Inclusion and Exclusion Criteria

The inclusion criteria of participants included: (1) age ≥18 years; (2) diagnosis of malignant tumors by a pathological or cytological method; (3) pain that met the cancer pain diagnostic criteria according to National Comprehensive Cancer Network Guidelines and was moderate to severe (NRS ≥4); (4) ability of patients or their families to read Chinese and use WeChat; (5) a normal verbal ability and performance status; and (6) agreement to participate in the study and sign the informed consent form.

Exclusion criteria of participants were: (1) nonmalignant pain; (2) hepatic dysfunction (alanine aminotransferase ≥2.5×upper limits of normal [ULN], aspartate aminotransferase ≥2.5×ULN, or total bilirubin ≥1.5×ULN) or renal dysfunction (serum creatinine ≥2.5×ULN); (3) participation in other clinical trials; and (4) hospitalization during the 4-week trial period.

Patient Enrollment and Intervention

Patient Enrollment

At the patients' first visit to the ambulatory clinic, the physician provided a detailed consultation and, then, determined a treatment plan after discussion with the pharmacist; an account was created for eligible patients. After registration, the pharmacist demonstrated the use of each MediHK module to patients, including what information was collected in each form and how to fill it out and how to switch the interface to send a form or question. Even though the operation of MediHK was simple enough, the training process was approximately 10 minutes. The specific time depended on patient understanding, ability, and proficiency in WeChat. After training, patients were assigned to a pain management team and were required to complete PROMs and forms. Patients at home could contact their pain management team at any time through MediHK if they had any trouble with pain. The pain management team was required to complete standardized clinical pain management training and had at least 10 years of hospital work experience for clinical pharmaceuticals for cancer pain before performing pain-management work.

Patients were observed for 4 weeks. At week 4, the patients were required to complete and submit the PROMs through MediHK or report through phone calls within 1 day. Patients could continue to use MediHK after the completion of the study.

Outcome Evaluation

The primary outcomes included pain intensity and pain interference in daily life. Secondary outcomes included PROMs, medication adherence, ADRs, and rehospitalization rates.

Statistical Analysis

All data were analyzed using SPSS software (version 22.0; IBM Corp), and all charts were made by the graphing software GraphPad (version 8.0.2 (263); GraphPad Prism). For measurement data, the normality test adopted the Kolmogorov-Smirnov method. If normally distributed, the data were expressed as mean (SD), and the comparison between the 2 groups used 2 independent sample *t* tests. If not normally distributed, the data were expressed as the median (IQR), and the comparison between groups underwent a Mann-Whitney U test. Counting data were expressed as a frequency and percentage. A chi-square (c^2) test or Fisher exact test was used for comparison between groups. We screened for factors affecting the pain intensity of outpatients with cancer pain by multivariate linear regression analysis (backward method, in=0.05, out=0.10). We defined $P<.05$ (test level=.05, two-tailed) as statistically significant.

Results

Principal Results

A total of 100 patients joined and completed this study, with 51 (51%) in the intervention group and 49 (49%) in the control group. Demographic information (ie, gender, age, height, and weight) and clinical information (ie, diagnosis, pain type, and site of pain) of the 2 groups were not statistically different, nor was the intensity, pain interference, and adherence to pain medications at baseline (*Ps*>.05; Table 1).

Table 1. Baseline characteristics of patients.

Variable	Intervention group (n=51)	Control group (n=49)	*P* value (statistical test)
Demographic information			
Male, n (%)	38 (75)	34 (69)	.57 (χ^2=0.325)
Age (years), mean (SD)	54.6 (14.0)	58.7 (14.8)	.16
Height (cm), median (IQR)	166.0 (160.0-170.0)	166.0 (160.0-169.5)	.29
Weight (kg), median (IQR)	55.0 (47.7-65.0)	55.0 (50.0-60.0)	.52
Diagnosis, n (%)			.33 (χ^2=4.653)
Lung cancer	16 (31)	24 (49)	
Gastrointestinal cancer	19 (37)	14 (29)	
Head and neck cancer	6 (12)	2 (4)	
Breast cancer	2 (4)	1 (2)	
Other	8 (16)	8 (16)	
Pain site, n (%)			.16 (χ^2=7.986)
≥2 sites	26 (51)	30 (61)	
Chest and abdomen	6 (12)	8 (16)	
Head and neck	6 (12)	5 (10)	
Back	6 (12)	5 (10)	
Limbs	5 (10)	0 (0)	
Other sites	2 (4)	1 (2)	
Pain type, n (%)			.08 (χ^2=6.801)
Mixed pain	20 (39)	21 (43)	
Visceral pain	27 (53)	16 (21)	
Neuropathic pain	3 (6)	9 (18)	
Body pain	1 (2)	3 (6)	
Pain intensity, median (IQR)[a]			
Worst pain	7 (5-8)	7 (6-9)	.16
Least pain	2 (1-3)	2 (1-3)	.26
Average pain	4 (2-6)	4 (3-6)	.33
Present pain	2 (1-4)	3 (1-4)	.17
Pain interference, median (IQR)[a]			
General activity	7 (4-10)	6 (3-8)	.31
Mood	5 (2-8)	5 (4-7)	.43
Walking ability	8 (2-10)	5 (2-9)	.27
Daily work	9 (4-10)	7 (4-9)	.10
Relationships	3 (2-6)	3 (2-6)	.94
Sleep	7 (4-9)	6 (5-9)	.88
Enjoyment of life	5 (2-7)	6 (2-8)	.61
Baseline adherence, n (%)			.10 (χ^2=2.784)
Nonadherence	3 (6)	8 (16)	
Incomplete adherence	26 (51)	25 (51)	
Complete adherence	22 (43)	16 (33)	

[a]These measures represent the baseline characteristics based on the Brief Pain Inventory.

BPI Outcomes

Pain intensity in the intervention group was significantly reduced compared with the control group. The worst pain scores, least pain scores, and average pain scores in the 2 groups were statistically different, with median values of 4 (IQR 3-7) vs 7 (IQR 5-8; *P*=.001), 1 (IQR 0-2) vs 2 (IQR 1-3; *P*=.02), and 2 (IQR 2-4) vs 4 (IQR 3-5; *P*=.001), respectively, favoring the intervention group. The difference in the present pain score of the 2 groups was not statistically significant (*P*=.23). However, the score of the intervention group was numerically lower than that of the control group (Table 2).

Table 2. Brief Pain Inventory outcomes at week 4.

BPI[a] item	Intervention group (n=51), median (IQR)	Control group (n=49), median (IQR)	*P* value
Pain intensity			
Worst pain	4 (3-7)	7 (5-8)	.001
Least pain	1 (0-2)	2 (1-3)	.02
Average pain	2 (2-4)	4 (3-5)	.001
Present pain	1 (0-3)	2 (0-4)	.23
Pain interference			
General activity	7 (4-8)	6 (3-8)	.76
Mood	3 (1-6)	4 (2-6)	.58
Walking ability	7 (4-10)	7 (3-8)	.32
Daily work	8 (6-10)	8 (6-9)	.15
Relationships	2 (1-4)	3 (1-5)	.64
Sleep	4 (1-7)	7 (3-8)	.10
Enjoyment of life	4 (2-7)	5 (2-8)	.43

[a]BPI: Brief Pain Inventory.

PROM Submission Through MediHK

The number of forms submitted by the intervention group patients was much higher than that of the control group (710 vs 95), with an average of 4.64 forms per person per day vs 0.06 forms per person per day, respectively. The most common forms submitted by the control group were the BPI (53/95, 56%), pain diary (17/85, 18%), and medication list (15/95, 16%; Table 3). Even though the control group patients did not receive reminders to fill out the forms, they still actively contacted the pain management team through MediHK due to uncontrollable pain intensity, interference in daily life, or severe ADRs.

Table 3. Number of PROMs submitted by the 2 groups.

Form	Intervention group (n=710), n (%)	Control group (n=95), n (%)	*P* value (χ^2=153.236)
Pain diary	495 (69.7)	17 (18)	<.001
ADR[a] form	87 (12.2)	7 (7)	<.001
BPI[b]	83 (11.7)	53 (56)	<.001
Medication list	31 (4.4)	15 (16)	<.001
MMAM[c]	14 (2.0)	3 (3)	<.001

[a]ADR: adverse drug reactions.

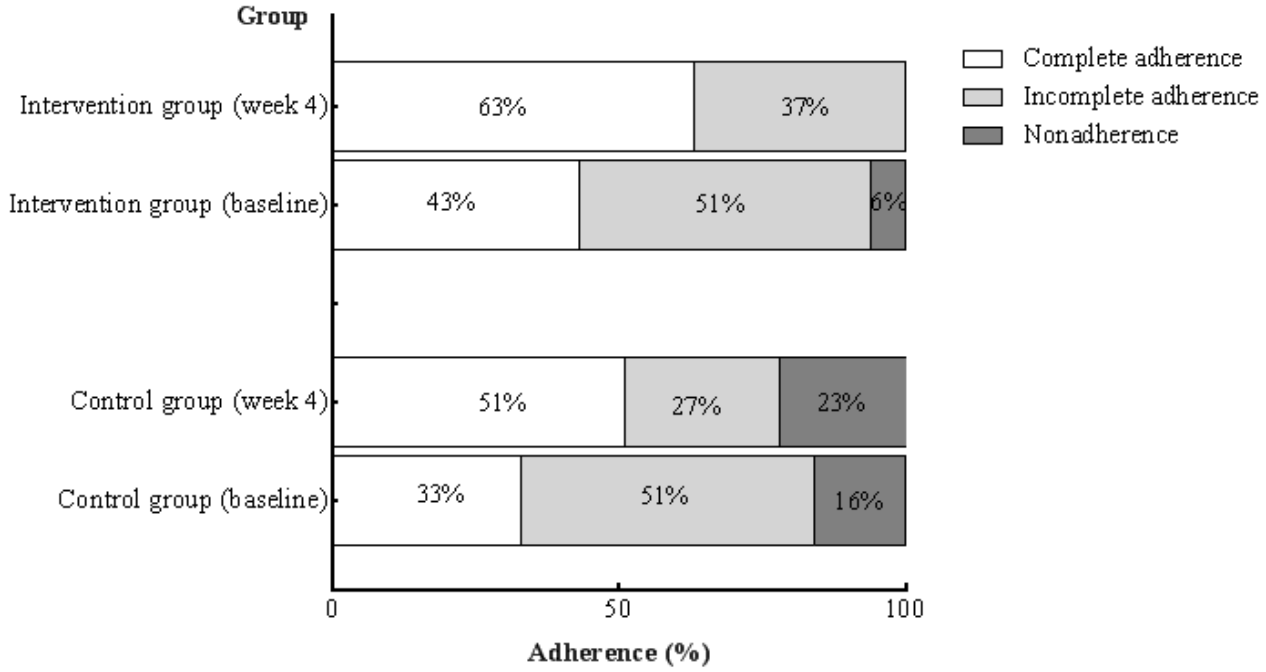
[b]BPI: Brief Pain Inventory.

[c]MMAM: Morisky Medication Adherence Measure.

Medication Adherence

The complete adherence rate in the intervention group increased from 43% (22/51) to 63% (32/51), while that of the control group increased from 33% (16/49) to 51% (25/49; χ^2=12.864; *P*<.001; Figure 1).

Figure 1. The adherence rate of the 2 groups at baseline and at week 4.



Adverse Drug Reactions

The overall incidence of ADRs was 36% (36/100) across the 2 groups at baseline and increased to 56% (56/100) at week 4. ADR incidence in the intervention group was significantly higher than in the control group (χ^2=8.990; *P*=.003). In addition, 3 cases of intestinal obstruction and 2 cases of delirium were observed in the intervention group. Table 4 shows the distribution of ADRs between groups.

Table 4. Adverse drug reactions between groups over 4 weeks.

Variable	Baseline			Week 4		
	Intervention group (n=51), n (%)	Control group (n=49), n (%)	*P* value (statistical test)	Intervention group (n=51), n (%)	Control group (n=49), n (%)	*P* value (statistical test)
Patients with ADR[a]	24 (47)	12 (25)	.02 (χ^2=5.525)	36 (71)	20 (41)	.003 (χ^2=8.990)
ADR type						
Constipation	18 (35)	6 (12)	—[b]	29 (57)	14 (29)	—
Nausea and vomiting	13 (26)	6 (12)	—	17 (33)	11 (22)	—
Drowsiness	4 (8)	3 (6)	—	9 (18)	3 (6)	—
Dizziness	6 (12)	1 (2)	—	9 (18)	3 (6)	—
Pruritus	2 (4)	—	—	3 (6)	—	—
Urinary retention	2 (4)	—	—	5 (10)	—	—
Ileus	—	—	—	3 (6)	—	—
Delirium	—	—	—	2 (2)	—	—

[a]ADR: adverse drug reaction.

[b]Not available.

Rehospitalization Rates During the 4 Weeks

The 2 groups had a similar rehospitalization rates within the 4-week trial. There was no significant difference between the 2 groups within 4 weeks (χ^2=0.010; *P*=.92).

Analysis of Pain Factors

We introduced possible factors that could contribute to pain intensity for each pain item in a multivariate linear regression analysis. The physician-pharmacist intervention through MediHK was an independent influencing factor for the most severe pain (β=–1.413; *P*=.005; Multimedia Appendix 8) and

average pain (β=–1.154; *P*=.003; Multimedia Appendix 9). Aside from medication adherence, the intervention was significantly related to the least pain (β=–.701; *P*=.02; Multimedia Appendix 10). No factors significantly influenced the present pain (gender β=1.078, *P*=.16; age β=.018, *P*=.26; height β=.063, *P*=.18; weight β=–.017, *P*=.46; adherence β=–.282, *P*=.40; intervention β=–.598, *P*=.17; Multimedia Appendix 11).

Discussion

Principal Results

The self-management of cancer pain is full of challenges, especially, for ambulatory patients. Approximately 70% to 90% of cancer patients can relieve pain adequately when carefully following the treatment guidelines. Now more fully developed, digital health helps to achieve good pain management in daily practice for ambulatory patients with cancer pain, particularly, in remote areas of China [6,11]. Patients' various demands in supporting self-management help encourage the development of a multimodal web application [16,17].

This study included a joint physician-pharmacist team that managed ambulatory patients with cancer pain through a WeChat-supported platform, MediHK, with promising results. Even if the control group did not receive a reminder to fill out the forms, patients in this group actively contacted the pain management team through MediHK to determine whether the medication plan needed to be adjusted due to either uncontrollable pain intensity, interference on daily life, or severe ADRs. The results revealed the patients' need to contact the professional team via MediHK for better pain management. The patients in the intervention group reported more ADRs compared with control group patients, primarily, because there were more reports obtained from intervention group patients. More ADRs were not in conflict with improving pain. For example, the pain management team added new drugs for pain treatment, which may have caused some ADRs, but most of them were tolerated after a few days and monitored closely by the pharmacist.

Comparison With Prior Work

Yang et al [18] developed an app named Pain Guard for better pain management of discharged patients. Its functions were similar to MediHK, such as self-evaluation, real-time medication consultation, and record-keeping. The differences were that, for MediHK, we combined the NRS, Face Pain Scale, and Verbal Rating Scale to assess pain intensity accurately, while Pain Guard had only the traditional scale, NRS. We designed the module to record more medication-related details from patients, including drug name, dose, frequency, initial stop time, pain relief after medication, and adverse reactions. In addition, the physician or pharmacist could send forms embedded in MediHK, such as the ADR or adherence assessment form, to patients according to their status. These functions were unavailable in the Pain Guard. Scriven et al [19] used the BPI to evaluate patients' pain while participating in the multisite telehealth group model. They found positive changes on the interference scale at the individual level (14% of patients) but no change at the group level. Another study offered standardized education and telemonitored for pain improvement, and BPI results indicated that, at 1 week, there were improvements in both the worst pain (from 7.3 to 5.7; *P*<.01) and average pain (from 4.6 to 3.8; *P*<.01) [20]. However, the portion of average pain rated ≥4 did not improve significantly because of the short study period [20]. Compared with telehealth, MediHK was more capable of real-time feedback.

Furthermore, we received more positive results because of our 4-week observation time. One study evaluated the effectiveness of pain management of a mobile phone app. Results showed that the pain relief rate was significantly different between the trial and the control groups (median 50, IQR 45-63 and median 0, IQR 0-25) [18]. Similarly, Sun et al [21] found a significant difference in the average pain score through an intelligent pain management system (mean 2.5, SD 0.42 vs mean 2.8, SD 0.47; *P*<.01). These findings support our vision of making full use of prescient and promising internet platforms to manage cancer pain. In another study with an internet application consisting of a pain diary and a pain education and consultation module, the present pain intensity and the worst pain intensity of patients in the intervention group were significantly reduced within 6 weeks [6]. MediHK found similar positive results and included more details in the pain diary module, such as recording all patient medication information and pain self-assessments at all times. It is worth mentioning that these studies were based on the NRS for pain assessment. However, MediHK also embedded the BPI form to consider pain itself and the interference it caused. During the 4 weeks, the worst pain intensity, least pain intensity, and average pain intensity of intervention group patients significantly reduced, with an average decrease of 1-2 points. In terms of pain interference, the impact of pain on patients' general activity, mood, relationships with others, and interests reduced. However, the difference was not statistically significant when compared with the control group. Intervention group patients showed significant improvements in adherence after 4 weeks, resulting from active interventions, raised awareness of patients, and real-time monitoring of ADRs, which was more accessible in the home setting through MediHK.

Regrettably, we did not record the impact of education status and age in keeping medical records. Patients who had never received education may take longer to keep records. However, since the included patients or their families were all able to use WeChat proficiently, we believed that MediHK was feasible; for patients who were too old or unable to record, family members or caregivers would help to send the form. In total, we accounted for the universal applicability of MediHK when we developed it to ensure easy operation. The only complicated step was the switch between the interface. However, this was emphasized when training in the outpatient clinic.

Knowledge deficits, inadequate pain assessment, misconceptions of pain, complex environments, and infrequent communication with health care providers are barriers in pain management. A joint physician-pharmacist team operating through a digital health platform can improve it. The cancer patient pain assessment was complicated. It is necessary to select quantification tools and assess the cause, location, quality, and relieving or aggravating factors of the pain comprehensively. The time that physicians spend on each patient is limited, and it is difficult to provide long-term and continuous monitoring.

The digital platform can better solve these problems. The platform trains patients to record their pain conditions in a more standardized and targeted manner. During clinical encounters, clinicians can spend more time addressing patients' concerns in a meaningful way, rather than running through checklists of questions [22]. In addition, this platform promotes patient self-management. It allows patients to pay attention to the daily changes in pain and offers a digital tool to seek out the help of a professional team when suffering from an intractable pain or serious ADRs.

It is essential to consider the clinical workflow, security and liability, and the time-cost. We conducted a preliminary investigation and consideration in the early stage and carried out several rounds of related process optimizations and software improvements. In addition, when patients first visited the clinic, we would state that their physicians and pharmacists would provide the home services via the platform, and patients trusted this service. Finally, patients signed an exemption agreement and informed consent to ensure medical safety before using the platform. The satisfaction of patients and medical workers on such digital health platforms matters. One study designed a module in a mobile app to survey overall satisfaction, and the questionnaire was completed by participants at the end of the study [18]. Another study also assessed patient satisfaction about the convenience and helpfulness of using mobile systems, receiving technical software support, receiving consultant and training courses, and prompt responses for help; the results indicated that patients had a high level of satisfaction toward these kinds of digital tools [21]. Our preliminary idea was to evaluate satisfaction by embedding a questionnaire. For patients, this included assessing the pattern of the platform and the pain management team's joint management, the content of medication education, the acceptability of response time, and the overall services. For the pain management team, this included evaluating the ease of operation of the platform, the acceptability of clinical workflow interference, and working-time costs. The questionnaire could contain an open-ended question, in which both patients and the pain management team are encouraged to provide suggestions regarding improvements to MediHK.

Study Strengths

There were some strengths of this study. First, this study was a real-world randomized controlled trial conducted in a large ambulatory clinic of a tertiary hospital. All patients were clinically recruited and randomly assigned. The integration of PROMs has not been a feature of other eHealth and (web) application–related studies, allowing this digital health study to help advance this field. In addition, real‐time reporting can facilitate just‐in‐time interventions based on an individual's current circumstance or environment. This study achieved real-time communication between ambulatory patients with cancer pain and health care providers through MediHK, extending medical services to ambulatory patients as a pathway for the self-management in home settings.

Study Limitations

The study had the several limitations. First, this study had abnormally high participation, which will not necessarily reflect what would happen when patients use the platform independently, because pharmacists would send daily notifications. Second, it was prospectively powered and conducted in a randomized manner, but inevitable confounding factors can exist in the real world. Multivariate linear regression can only explain a small part of the influence of different pain intensity types. Third, since the study was conducted in a single tertiary hospital, applying this approach in other clinical settings may require some individualization to meet specific needs. Fourth, the observation time of only 4 weeks limited the long-term application of the results. Fifth, this study lacked further assessment about buy-in from both patients and the pain management team.

Conclusions

The joint physician-pharmacist team operating through MediHK enhanced communication, optimized outcomes, and promoted self-management of patients in home settings. This study supports the feasibility of integrating the internet into patient self-management of cancer pain. In the future, it will be necessary to enlarge the sample size to further explore the long-term effects of this method on the self-management of ambulatory patients with cancer pain.

Acknowledgments

MediHK was developed by the joint efforts of the College of Mathematics and Computer Science, Hunan Normal University and the Department of Pharmacy of Xiangya Hospital, Central South University. The authors sincerely thank all of the patients and their relatives for cooperating in this study. We thank all physicians for their support and participation as well as the College of Mathematics and Computer Science, Hunan Normal University for their subsequent services.

The study was supported by Changsha Natural Science Foundation of 2020 (kq2007039), National Clinical Research Center for Geriatric Disorders, Xiangya Hospital (XYYYJSTG-15), Parallel and Distributed Processing for the Stable Support Project of the National Defense Science and Technology Key Laboratory (WDZC20205500121), The Fundamental Research Funds for the Central Universities of Central South University (2021zzts1040), and Science and Technology Development Center of Chinese Pharmaceutical Association (CMEI2021KPYJ (JZYY) 00301).

Authors' Contributions

HLM, K-KL, S-SS, X-PC, and JX designed the study. K-KL, LZ, and HLM wrote the manuscript. K-KL, HLM, H-XH, and LZ performed the research. JX and H-XH analyzed the data. Y-MH, YC, W-HL, and F-ZL contributed new reagents and analytical tools. F-ZL provided technology support.

Conflicts of Interest

HLM is on the board of directors for Vyant Bio. He is one of the founders of Clariifi LLC and a consultant to Viecure and eviCORE Health Solutions.

Multimedia Appendix 1
Screenshots of the patient interface, including registration, homepage, and reminders.
[PNG File , 3193 KB - mhealth_v9i8e24555_app1.png]

Multimedia Appendix 2
Screenshots of the medical interface, including reminders, review, homepage, and message-sending.
[PNG File , 2674 KB - mhealth_v9i8e24555_app2.png]

Multimedia Appendix 3
Screenshots of the introduction of the MediHK.
[PNG File , 2422 KB - mhealth_v9i8e24555_app3.png]

Multimedia Appendix 4
Medication history list. English translation shown on right.
[PNG File , 267 KB - mhealth_v9i8e24555_app4.png]

Multimedia Appendix 5
Adherence Assessment Measurement. English translation shown on right.
[PNG File , 1605 KB - mhealth_v9i8e24555_app5.png]

Multimedia Appendix 6
Adverse Reaction Form. English translation shown on right.
[PNG File , 1917 KB - mhealth_v9i8e24555_app6.png]

Multimedia Appendix 7
Pain diary. English translation shown on right.
[PNG File , 2832 KB - mhealth_v9i8e24555_app7.png]

Multimedia Appendix 8
The independent factors influencing worst pain intensity.
[DOC File , 26 KB - mhealth_v9i8e24555_app8.doc]

Multimedia Appendix 9
The independent factors influencing average pain intensity.
[DOC File , 26 KB - mhealth_v9i8e24555_app9.doc]

Multimedia Appendix 10
The independent factors influencing least pain intensity.
[DOC File , 25 KB - mhealth_v9i8e24555_app10.doc]

Multimedia Appendix 11
The independent factors influencing present pain intensity.
[DOC File , 25 KB - mhealth_v9i8e24555_app11.doc]

Multimedia Appendix 12
CONSORT-eHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 2350 KB - mhealth_v9i8e24555_app12.pdf]

References

1. Neufeld NJ, Elnahal SM, Alvarez RH. Cancer pain: a review of epidemiology, clinical quality and value impact. Future Oncol 2017 Apr;13(9):833-841. [doi: 10.2217/fon-2016-0423] [Medline: 27875910]
2. Enting RH, Oldenmenger WH, Van Gool AR, van der Rijt CCD, Sillevis Smitt PAE. The effects of analgesic prescription and patient adherence on pain in a dutch outpatient cancer population. J Pain Symptom Manage 2007 Nov;34(5):523-531 [FREE Full text] [doi: 10.1016/j.jpainsymman.2007.01.007] [Medline: 17664055]
3. Fairchild A. Under-treatment of cancer pain. Curr Opin Support Palliat Care 2010 Mar;4(1):11-15. [doi: 10.1097/SPC.0b013e328336289c] [Medline: 20040878]
4. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. CA Cancer J Clin 2016;66(2):115-132 [FREE Full text] [doi: 10.3322/caac.21338] [Medline: 26808342]
5. Oldenmenger WH, Sillevis Smitt PAE, van Dooren S, Stoter G, van der Rijt CCD. A systematic review on barriers hindering adequate cancer pain management and interventions to reduce them: a critical appraisal. Eur J Cancer 2009 May;45(8):1370-1380. [doi: 10.1016/j.ejca.2009.01.007] [Medline: 19201599]
6. Oldenmenger WH, Baan MAG, van der Rijt CCD. Development and feasibility of a web application to monitor patients' cancer-related pain. Support Care Cancer 2018 Feb;26(2):635-642 [FREE Full text] [doi: 10.1007/s00520-017-3877-3] [Medline: 28929433]
7. Persell SD, Eder M, Friesema E, Connor C, Rademaker A, French DD, et al. EHR-based medication support and nurse-led medication therapy management: rationale and design for a three-arm clinic randomized trial. J Am Heart Assoc 2013 Oct 24;2(5):e000311 [FREE Full text] [doi: 10.1161/JAHA.113.000311] [Medline: 24157649]
8. Bennett AV, Jensen RE, Basch E. Electronic patient-reported outcome systems in oncology clinical practice. CA Cancer J Clin 2012;62(5):337-347 [FREE Full text] [doi: 10.3322/caac.21150] [Medline: 22811342]
9. Kim CH, Chung CK, Choi Y, Shin H, Woo JW, Kim S, et al. The usefulness of a mobile device-based system for patient-reported outcomes in a spine outpatient clinic. Spine J 2016 Jul;16(7):843-850. [doi: 10.1016/j.spinee.2016.02.048] [Medline: 26961200]
10. Wagner LI, Schink J, Bass M, Patel S, Diaz MV, Rothrock N, et al. Bringing PROMIS to practice: brief and precise symptom screening in ambulatory cancer care. Cancer 2015 Mar 15;121(6):927-934 [FREE Full text] [doi: 10.1002/cncr.29104] [Medline: 25376427]
11. Hochstenbach LMJ, Courtens AM, Zwakhalen SMG, Vermeulen J, van Kleef M, de Witte LP. Co-creative development of an eHealth nursing intervention: Self-management support for outpatients with cancer pain. Appl Nurs Res 2017 Aug;36:1-8. [doi: 10.1016/j.apnr.2017.03.004] [Medline: 28720227]
12. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. J Adv Nurs 2000 Oct;32(4):1008-1015. [Medline: 11095242]
13. Wang XS, Mendoza TR, Gao SZ, Cleeland CS. The Chinese version of the Brief Pain Inventory (BPI-C): its development and use in a study of cancer pain. Pain 1996 Oct;67(2-3):407-416. [doi: 10.1016/0304-3959(96)03147-8] [Medline: 8951936]
14. Morisky DE, Green LW, Levine DM. Concurrent and predictive validity of a self-reported measure of medication adherence. Med Care 1986 Jan;24(1):67-74. [Medline: 3945130]
15. Tzeng JI, Chang C, Chang H, Lin C. Assessing analgesic regimen adherence with the Morisky Medication Adherence Measure for Taiwanese patients with cancer pain. J Pain Symptom Manage 2008 Aug;36(2):157-166 [FREE Full text] [doi: 10.1016/j.jpainsymman.2007.10.015] [Medline: 18411015]
16. van Houtum L, Rijken M, Heijmans M, Groenewegen P. Self-management support needs of patients with chronic illness: do needs for support differ according to the course of illness? Patient Educ Couns 2013 Dec;93(3):626-632. [doi: 10.1016/j.pec.2013.08.021] [Medline: 24029582]
17. Hochstenbach LMJ, Zwakhalen SMG, Courtens AM, van Kleef M, de Witte LP. Feasibility of a mobile and web-based intervention to support self-management in outpatients with cancer pain. Eur J Oncol Nurs 2016 Aug;23:97-105 [FREE Full text] [doi: 10.1016/j.ejon.2016.03.009] [Medline: 27456381]
18. Yang J, Weng L, Chen Z, Cai H, Lin X, Hu Z, et al. Development and Testing of a Mobile App for Pain Management Among Cancer Patients Discharged From Hospital Treatment: Randomized Controlled Trial. JMIR Mhealth Uhealth 2019 May 29;7(5):e12542 [FREE Full text] [doi: 10.2196/12542] [Medline: 31144672]
19. Scriven H, Doherty DP, Ward EC. Evaluation of a multisite telehealth group model for persistent pain management for rural/remote participants. Rural Remote Health 2019 Mar;19(1):4710 [FREE Full text] [doi: 10.22605/RRH4710] [Medline: 30884951]
20. Kim HS, Shin SJ, Kim SC, An S, Rha SY, Ahn JB, et al. Randomized controlled trial of standardized education and telemonitoring for pain in outpatients with advanced solid tumors. Support Care Cancer 2013 Jun;21(6):1751-1759 [FREE Full text] [doi: 10.1007/s00520-013-1722-x] [Medline: 23338230]
21. Sun Y, Jiang F, Gu JJ, Wang YK, Hua H, Li J, et al. Development and Testing of an Intelligent Pain Management System (IPMS) on Mobile Phones Through a Randomized Trial Among Chinese Cancer Patients: A New Approach in Cancer Pain

Management. JMIR Mhealth Uhealth 2017 Jul 25;5(7):e108 [FREE Full text] [doi: 10.2196/mhealth.7178] [Medline: 28743681]
22. Jim HSL, Hoogland AI, Brownstein NC, Barata A, Dicker AP, Knoop H, et al. Innovations in research and clinical care using patient-generated health data. CA Cancer J Clin 2020 May;70(3):182-199 [FREE Full text] [doi: 10.3322/caac.21608] [Medline: 32311776]

Abbreviations

ADR: adverse drug reaction
BPI: Brief Pain Inventory
MediHK: Medication Housekeeper
NRS: Numerical Rating Scale
PROM: patient-reported outcome measures
ULN: upper limits of normal

Edited by L Buis; submitted 27.09.20; peer-reviewed by D Lin, J Vangara, F Zheng; comments to author 02.04.21; revised version received 01.05.21; accepted 15.07.21; published 16.08.21.

Please cite as:
Zhang L, McLeod HL, Liu KK, Liu WH, Huang HX, Huang YM, Sun SS, Chen XP, Chen Y, Liu FZ, Xiao J
Effect of Physician-Pharmacist Participation in the Management of Ambulatory Cancer Pain Through a Digital Health Platform: Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(8):e24555
URL: https://mhealth.jmir.org/2021/8/e24555
doi:10.2196/24555
PMID:34398796

Original Paper

Relationship Between Perceived Risks of Using mHealth Applications and the Intention to Use Them Among Older Adults in the Netherlands: Cross-sectional Study

Nicky Sabine Klaver[1], MSc; Joris van de Klundert[1,2], MSc, PhD; Roy Johannes Gerardus Maria van den Broek[1], MSc; Marjan Askari[1], MSc, PhD

[1]Erasmus School of Health Policy & Management, Erasmus University, Rotterdam, Netherlands

[2]Prince Mohammad Bin Salman College of Business & Entrepreneurship, King Abdullah Economic City, Saudi Arabia

Corresponding Author:
Marjan Askari, MSc, PhD
Erasmus School of Health Policy & Management
Erasmus University
PO Box 1738
Rotterdam, 3000 DR
Netherlands
Phone: 31 104088641
Email: m.askari@askari.nl

Abstract

Background: Considering the increasing demand for health services by older people and the ongoing COVID-19 pandemic, digital health is commonly viewed to offer a pathway to provide safe and affordable health services for older adults, thus enabling self-management of their health while health care systems are struggling. However, several factors cause older people to be particularly reluctant to adopt digital health technologies such as mobile health (mHealth) tools. In addition to previously studied technology acceptance factors, those related to perceived risks of mHealth use (eg, leakage of sensitive information or receiving incorrect health recommendations) may further diminish mHealth adoption by older adults.

Objective: The aim of this study was to explore the relationship between perceived risks of using mHealth applications and the intention to use these applications among older adults.

Methods: We designed a cross-sectional study wherein a questionnaire was used to collect data from participants aged 65 years and older in the Netherlands. Perceived risk was divided into four constructs: privacy risk, performance risk, legal concern, and trust. Linear regression analyses were performed to determine the associations between these perceived risk constructs and the intention to use mHealth applications.

Results: Linear regression per perceived risk factor showed that each of the four constructs is significantly associated with the intention to use mobile medical applications among older adults (adjusted for age, sex, education, and health status). Performance risk (β=–.266; *P*=<.001), legal concern (β=–.125; *P*=.007), and privacy risk (β=–.100; *P*=.03) were found to be negatively correlated to intention to use mHealth applications, whereas trust (β=.352; *P*=<.001) was found to be positively correlated to the intention to use mHealth applications.

Conclusions: Performance risk, legal concern, and privacy risk as perceived by older adults may substantially and significantly decrease their intention to use mHealth applications. Trust may significantly and positively affect this intention. Health care professionals, designers of mHealth applications, and policy makers can use these findings to diminish performance risks, and tailor campaigns and applications to address legal and privacy concerns and promote mHealth uptake and health care access for older adults, especially during the COVID-19 pandemic.

(JMIR Mhealth Uhealth 2021;9(8):e26845) doi:10.2196/26845

KEYWORDS

mHealth; older adults; perceived risks; intention to use; adoption; covid-19; digital health

Introduction

Due to the shift in age distribution, an increasingly larger proportion of the world population will consist of adults over 65 years of age [1-3]. As we grow older, our demand for health care increases [4,5]. The global increase in the proportion of older adults is therefore expected to significantly increase global health care usage and costs [6]. To keep the health care system sustainable in the future, governments and societies are considering technology as a promising solution for health service delivery innovation and for service expansion without increasing human resource capacity [7-9].

Moreover, the COVID-19 outbreak has had extensive consequences for the provision of health care. This holds especially for older patients whose health care usage is likely to be higher while the pandemic makes it more critical for them to stay home, as age significantly determines the clinical features and prognosis of COVID-19 [10-12]. Through their rapid uptake, mobile health (mHealth) technologies enable health care without the need for face-to-face contact [10,13,14]. mHealth includes health-supporting applications on wireless devices such as tablets or smartphones [15-17]. These applications can assist independently living older adults, for instance, by monitoring clinical signs, collecting health information, or promoting a healthy lifestyle [18-21]. mHealth has shown to be able to improve care, self-management, and self-efficacy, as well as promote better behavior and medication adherence of older adults [14,19,22,23]. Unfortunately, however, older adults are less likely to adopt new technologies such as mHealth [24-28]. Before the pandemic, almost 50% of Dutch older adults had no intention to use mHealth applications [29].

Older people are deserving of special attention when it comes to technology adoption because of their different attitude toward technology [30,31]. For instance, they are more likely to perceive risks in the adoption of new technologies [32,33]. Potential risks, such as sensitive information leakage or incorrect health recommendations, may keep older people from using mHealth applications even during the COVID-19 pandemic when they are deemed especially beneficial. Earlier studies have shown that perception of risks are important barriers for mHealth acceptance and adoption [34-39]. Nevertheless, few studies have addressed the risks perceived by older people in relation to overall mHealth adoption or their intention to use such applications, which predominantly determines their adoption [40,41].

Deng et al [35] found that the intention to adopt mHealth is influenced by the perceived risks of using it among the general population. Perceived risks are one's perception of uncertainty in the use of mHealth and include trust, performance risk, legal concern, and privacy risk. However, their study included very few respondents—those aged 65 years and older. Moreover, they focused on the hospital context where the use of mHealth applications is often limited to web-based consultation with a practitioner [35]. As a result, aging people who are not hospitalized remain an under-researched population concerning the perceived risks of mHealth adoption. Advancing the understanding of mHealth adoption among independently living older adults, therefore, has general importance because of their health care utilization and is especially relevant during the COVID-19 pandemic.

This study focuses on the perceived risks of mHealth adoption among independently living older adults. More specifically, the research aim is to determine the relationship between the risks of using mHealth applications as perceived by independently living older adults and their intention to use these applications. We set out to assess the validity and significance of perceived risks factors determining the intention to use the medical applications in a quantitative study involving a large sample of independently living older adults in the Netherlands.

Methods

Overview

The technology acceptance model (TAM) was developed to analyze the usage behavior of information technologies in organizational contexts [42]. TAM suggests that actual usage behavior is driven by the behavioral intention to use a system, as is also the case for the subsequent unified theory of acceptance and use of technology (UTAUT) model [43], and this has been empirically confirmed for mHealth usage among older adults [40,41,43]. Therefore, in this study, we selected behavioral intention to use mHealth applications from TAM as an outcome of interest. Building on the work of Askari et al [29], the statements operationalizing intention to use are taken from Venkatesh and Davis [44] and translated into Dutch, adding one new statement to account for linguistic differences.

Perceived Risk

Perceived risk, defined by Deng et al [35] as "one's perception of uncertainty in the use of mHealth services and its severity in terms of consequences," is measured with four constructs: privacy risk, performance risk, legal concern, and trust. In a 2018 empirical study, these authors identified an association between privacy risk, performance risk, legal concern, and trust with the intention to use mHealth in the general Chinese population. In this study, we adopted the statements, operationalizing these constructs from Deng et al [35] and translated them into Dutch to fit the context of independently living Dutch older adults. The corresponding English-language statements used to measure these constructs, as well as the questions to measure intention to use, are presented in Multimedia Appendix 1 [35,44]. Below, we explain the four different constructs that together capture perceived risk.

Privacy Risk

In this study, *privacy risk* refers to the extent to which an individual believes personal information abuse may occur because of mHealth application usage [35]. Previous studies on older adults have identified privacy concerns as a barrier to adopt health care technologies [45-47]. In our research context, we hypothesized that privacy concerns of older adults are negatively associated with their behavioral intention to use mHealth. We expect that older adults seek to have control over their lives as much as possible since perceived control is identified as an important factor in the well-being of aging people [48]. Sharing sensitive information over the internet may

be perceived to diminish perceived control. Furthermore, older adults may be less familiar with technologies and, therefore, insecure about future destinations of personal information shared through an mHealth application.

Performance Risk

Performance risk is defined as the extent to which an individual doubts the capability of mHealth applications to realize desired outcomes [35]. Many older adults do not feel that they are able to use smartphones or tablets properly [30,49]. This could make them question the suitability of mHealth applications to manage their health. Moreover, many older people are skeptical that technology will replace health care professionals [45]. This also suggests they may question the quality and usability of health care technologies in comparison to traditional healthcare services. Therefore, we hypothesized that older adults are less likely to have a behavioral intention to use mHealth applications when perceiving performance risks.

Legal Concern

Legal concern refers to an individual's worries regarding inappropriate law enforcement for mHealth applications [35]. For instance, older adults may prefer mHealth applications provided by third parties rather than by health care providers they visit, to prevent personal data to be illegally combined with clinical data. As our study targets mHealth applications to be used from home, the data protection is not covered by legislation applying to inpatient settings. Therefore, legal concern is hypothesized to be negatively associated with the intention to use mHealth applications. Our hypothesis is further corroborated by the lack of clarity about which laws cover mHealth disputes [50].

Trust

Finally, *trust* is defined as the perceived credibility of an mHealth application and the people behind it [35]. Older adults are less likely to trust assistive health care technologies [45]. Previous studies on the general population reveal a positive association between trust and the intention to use mHealth technologies [35,51-53]. We hypothesized the same association to hold for older adults.

Study Design and Data Collection

A cross-sectional study was designed for this research, in which older adults over the age of 65 years were approached from February through June 2020. We developed a questionnaire and administered it digitally during the initial months of the COVID-19 pandemic. Data were collected by four data assistants in cooperation with different organizations, such as living facilities, senior citizen associations, and health service provider organizations, and via different web-based channels and mailing lists. The data has mainly been collected in the regions of Noord-Brabant, Utrecht, and Zuid-Holland. We have not been able to keep track of the number of recipients of the distributed web-based questionnaire as the cooperation organizations reached out to their clients and members for us. The number of completed questionnaires is reported in the Results section. The reporting of the web-based questionnaire follows the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) checklist [54] and is presented in Multimedia Appendix 2.

The inclusion criteria for participants were as follows:

- The participant is 65 years of age or older
- The participant does not have cognitive impairments
- The participant lives independently

All respondents were asked to sign an informed consent form before participation. Thereafter, they could answer the questionnaire anonymously. The purpose of the project and information about the questionnaire, data management, and privacy of the participant were provided at the start of the questionnaire. Assistance and explanations were provided to participants who needed help filling out the questionnaire via telephone or email when requested. Data assistants entered the completed questionnaires into an SPSS database (IBM Corp) and pseudonymized the data to ensure anonymity.

Statistical Analyses

Privacy Constructs and the Dependent Variable

The constructs privacy risk, performance risk, legal concern, and trust were adopted from a validated instrument and designed on a 5-point Likert scale, ranging from 1 = "strongly disagree" to 5 = "strongly agree" [35]. Per perceived risk construct, a score was computed by calculating the average score of all the related statements of that construct. These average scores acted as the independent variables in the analysis. The dependent variable *intention to use mHealth* was similarly calculated [55]. Participants with one or more missing values at the intention to use mHealth statements were deleted from further analysis. To test the internal reliability and validate the reliability of the statements for Dutch older adults, we calculated the Cronbach α for each construct. A construct was considered as reliable if Cronbach α was greater than .70 [56,57]. As an additional reliability test, a correlation matrix was calculated to test if the perceived risks factors were independent. Interdependent factors were not included together in the subsequent regression analysis to avoid multicollinearity [58].

Univariate and Multivariate Linear Regression Analyses

The calculated perceived risk factor scores served as an input for the univariate and multivariate linear regression analyses to examine the relationship between (each of) the perceived risk factors as independent variables and intention to use mHealth as the dependent variable. Univariate linear regression analyses were used to summarize the linear relationship between each perceived risk construct and intention to use mHealth, without controlling. Standardized coefficients with 95% CIs and *P* values were reported. Multivariate linear regression analyses were performed to calculate the unstandardized and standardized coefficients with 95% CIs and *P* values, to evaluate the relationship between each perceived risk construct and intention to use mHealth while controlling for age, sex, education, and health status. The rationale for choosing these control variables is explained below.

Control Variables

We included the following control variables: age, sex, education level, and health status. A correlation matrix was developed to test whether the control variables were independent [58]. The variables *sex* and *education level* were recoded into dummy variables as these variables are categorical variables. *Age* has been reported to be negatively related to intention to use technology [25,59]. Moreover, the risk of user errors increases with age due to an increase in physical and perceptual difficulties [60]. *Sex* was included as a control variable because previous studies on older adults determined that men are more likely to use technologies than women [27,59,61]. Women are generally more concerned than men [62]. Moreover, *education level* was included as a control variable, as more highly educated people are more likely to use mHealth technologies [35,59,63]. Finally, we included *health status* as a control variable since older adults may perceive more benefit from using mHealth as their health status would be relatively worse [63]. On the other hand, older adults with poorer health status may consider themselves more vulnerable and, therefore, more sensitive to perceived risks and mistrust [64].

Validity and Reliability

Internal validity benefits from using validated instruments [35,42]. The questionnaire was validated by 5 experts (3 eHealth experts, 1 geriatric nurse, and 1 geriatrician). The usability and technical functionality of the web-based questionnaire was tested by 3 data assistants. The questionnaire is available on request. The database was checked for completeness and input errors by comparing a sample of paper questionnaires with the corresponding database information. To increase external validity, data collection took place in several different geographical locations within the Netherlands. Finally, Cronbach α was calculated for each factor to test reliability.

The study was approved by the Medical Ethical Committee of Erasmus Medical Center (number MEC-2018-120). The analysis was conducted using SPSS Statistics software (version 25; IBM Corp).

Results

Respondents' Characteristics

Our sample consisted of 481 respondents. However, 18 respondents were excluded from the total for not filling out the questions regarding the dependent variable. The respondents in this sample had a mean age of 74 (SD 5.77) years. Almost all respondents (456/463, 98.5%) had prior experience with using the internet, but only 127 (26.6%) had ever used medical applications before. Further details on the respondent characteristics can be found in Table 1.

Table 1. Baseline characteristics of the study cohort (N=463).

Characteristics	Participants, n (%)
Sex	
Male	231 (50.9)
Female	223 (49.1)
Age (years)	
65-74	270 (58.4)
75-84	164 (35.5)
≥85	28 (6.1)
Education level	
No or lower education	64 (13.9)
Intermediate education	232 (50.3)
Higher education	165 (35.8)
General health	
Poor	4 (0.9)
Fair	86 (18.6)
Good	215 (46.4)
Very good	108 (23.3)
Excellent	50 (10.8)
Prior experience with the internet	
Yes	456 (98.5)
No	7 (1.5)
Prior experience with mHealth	
Yes	127 (27.4)
No	336 (72.6)
Participation (data collection timepoint)	
February 2020	1 (0.2)
March 2020	108 (23.3)
April 2020	165 (35.6)
May 2020	188 (40.6)
June 2020	1 (0.2)

Cronbach Alpha and Correlation Analyses

Cronbach α scores, which show the internal consistency of the items within each acceptance factor, are shown in Table 2. The scores were well above the recommended limit of .70, indicating acceptable reliability [57].

The outcome of the correlation analysis is presented in Table 3. Perceived risk factors were found to be significantly and substantially correlated to one another. Therefore, the factors are not jointly included in the multivariate linear regression analysis. The control variables were also combined in a correlation analysis (see Multimedia Appendix 3). This analysis showed that the control variables were not substantially correlated to one another and could, therefore, be jointly included as control variables in the multivariate linear regression analysis.

Table 2. Cronbach α values for acceptance factors of the questionnaire.

Cluster (number of statements within a construct)	Cronbach α
Intention to use (n=3)	.959
Privacy risk (n=4)	.911
Performance risk (n=4)	.868
Legal concern (n=3)	.921
Trust (n=5)	.863

Table 3. Correlation analysis between the four risk factors.

Variable	Privacy risk	Performance risk	Legal concern	Trust
Privacy risk				
r	1	0.510	0.719	–0.202
P value (2-tailed)	__[a]	<.001	<.001	<.001
Performance risk				
r	0.510	1	.576	–0.323
P value (2-tailed)	<.001	—	<.001	<.001
Legal concern				
r	0.719	.576	1	–0.264
P value (2-tailed)	<.001	<.001	—	<.001
Trust				
r	–0.202	–0.323	–0.264	1
P value (2-tailed)	<.001	<.001	<.001	—

[a]Not applicable.

Univariate and Multivariate Analyses

The results of the univariate and multivariate regression analyses are summarized in Table 4. All factors were found to be significantly associated with intention to use mHealth. Privacy risk, performance risk, and legal concern were found to be negatively associated with intention to use, and trust was found to be positively associated with intention to use. In the multivariate regression analyses, we controlled for sex, age, education level, and health status. Privacy risk, performance risk, and legal concern continued to have a significantly negative coefficient. Performance risk had a negative coefficient of –0.266. This coefficient is large in comparison to the coefficients for privacy risk (–0.099) and legal concern (–0.124). Trust had the largest significant (positive) coefficient (0.350). The multivariate regression analyses (see Multimedia Appendix 3) also show that health status is significantly and negatively correlated to intention to use when considered a control variable for each of the four risk factors, whereas education was only significant in the model with trust. Age was significantly related to intention to use, except when being a control variable for legal concern. Sex was not found to be significant in any scenario.

Table 4. Results of univariate and multivariate linear regression analyses, and coefficients for perceived risk factors.

Variable	Univariate linear regression		Multivariate linear regression[a]			
	Standardized β coefficient (95% CI)	*P* value	Unstandardized coefficient B (95% CI)	Standardized β coefficient	*P* value	Adjusted R^2
Privacy risk	–.124 (–0.224 to –0.034)	.008	–0.103 (–0.195 to –0.011)	–.100	.03	0.089
Performance risk	–.331 (–0.537 to –0.315)	<.001	–0.337 (–0.450 to –0.225)	–.266	<.001	0.147
Legal concern	–.167 (–0.284 to –0.085)	<.001	–0.136 (–0.235 to –0.038)	–.125	.007	0.093
Trust	.375 (0.464 to 0.735)	<.001	0.555 (0.422 to 0.687)	.352	<.001	0.202

[a]Adjusted for age, sex, education, and health status.

Discussion

Principal Results and Comparison With Prior Work

This study explored whether perceived risks influence the intention to use mHealth applications among independently living older adults. Our findings showed that privacy risk, performance risk, legal concern, and trust were significantly associated with participants' intention to use mHealth applications. The perceived risks (ie, privacy risk, performance risk, legal concern) were negatively associated with intention to use, whereas *trust* was positively associated with intention to use.

Trust had the largest correlation coefficient and explained more than 20% of the variance in intention to use when the controls are added, indicating that it has a considerable positive effect on the behavioral intention to use mHealth. This finding indicates that older adults who perceive doctors accessible via mHealth applications as trustworthy and reliable have a higher intention to use these applications, as hypothesized in this study. These findings broadly confirm previous findings obtained for the general population in China and strengthen existing evidence that trust is an important determinant for the intention to use mHealth applications [35,51-53].

Performance risk was significantly and negatively associated with individuals' intention to use mHealth. These findings confirm our hypothesis stating that independently living older adults who doubt whether mHealth applications can meet their health care needs have a lower intention to use these applications. This further confirms previously reported findings for the general population and may be more valid for older adults, as they are more likely to fear that technologies will replace health care professionals [35,45].

Legal concern was negatively and significantly associated with intention to use. This confirms our hypothesis, which was based on the argument that older adults who are more likely to worry about inappropriate law enforcement are less likely to have a behavioral intention to use mHealth. The significance of the relationship differs from previous research in which the relationship between legal concern and intention to use was not found to be significant [35]. This may be explained by the fact that Deng et al [35] addressed the general Chinese population, in the hospital context, wherein mHealth use is more limited and legal issues may be less of a concern to respondents because of specific health laws being in place in this context. In addition, discussions on legal aspects of mHealth applications received considerable attention in the Netherlands during the COVID-19 outbreak, especially about a "Corona App" [65,66]. This may have raised concerns among independently living older adults. Legal concerns regarding mHealth use appear to have received little attention in the scientific literature and form a relevant area for further research.

Privacy risk also was significantly and negatively associated with participants' intention to use mHealth, albeit with a smaller coefficient. This finding confirms our hypothesis and is consistent with previous results reported by Deng et al [35] in the general Chinese population. Privacy risk is not directly linked to the functions of the mHealth app, as it involves the confidentiality of personal health information during the use of mHealth. Hence, compared with performance risk, privacy risk may likely exert less effect on the intention to use mHealth, which explains the smaller coefficient [35].

As the correlation analysis showed, the perceived risk factors are significantly and substantially correlated to one another. These results are not in accordance with the results from Deng et al [35], where a high discriminant validity was shown between the factors. This could be explained by the fact that Deng et al's 2018 study included very few respondents of 65 years and above, and older adults are more prone to perceive mHealth risks than younger people [67,68]. Another related explanation might be that there are underlying shared determinants of the risks perceived by older adults.

A recent study from the Netherlands showed that the adoption rate of a COVID-19 tracing app was significantly lower for older adults than for young adults [69]. One of their hypotheses for this lower adoption rate was that older adults would feel insufficiently protected by a contract tracing app. Older adults felt insufficiently protected because of the different perceived risk factors as shown in our study, thereby leading to a lower adoption rate.

Relationships Between Trust and Other Factors

Trust has been proposed to function as a mediator of five relationships between the three identified perceived risks (ie, privacy risk, performance risk, and legal concern) and the two TAM factors (perceived ease of use and perceived usefulness) with intention to Use [35]. Certainly, these hypotheses have intuitive appeal and, to some extent, theoretical support as well [35,37,38,41]. The empirical results of Deng et al [35] accept two of these five hypotheses and reject the other three.

The theoretical support for the relationship of these variables with trust clearly depends on the definition and operationalization of "trust." Following the construct definition presented by Deng et al [35], the operationalization adopted in this study focuses on trust in the medical doctors that the mHealth applications connects the user with. Such trust in medical doctors is essentially different from trust in the technology itself, and the intuitive and theoretical arguments cannot be assumed to remain unaffected. In fact, there is literature to support that the technology acceptance factors are influenced by trust in "entities behind the system" [70], suggesting that trust is a determinant of perceived ease of use and perceived usefulness rather than a mediator of their effect on the intention to use.

As shown in Multimedia Appendix 3, the two TAM variables perceived usefulness and perceived ease of use are indeed significantly related to trust in our study. However, in view of the definition and the arguments above, further appropriately designed research on the nature and direction of these relationship is called for.

Similar reflections are in place regarding the relationships between trust and the three perceived risk factors. Multimedia Appendix 3 presents univariate and multivariate regression analyses with trust as the dependent variable for the three risk factors. It shows that performance risk, privacy risk, and legal concern all are negatively and significantly associated with trust. This finding contrasts with previous findings by Deng et al [35] who report that performance risk and privacy risk were significantly and negatively associated with trust, whereas legal concern was not significantly related to trust. This difference in findings might well be explained by a general difference in legal concern between the study populations, which differ in culture, age, and state. In view of the definition of trust, we call for caution to state whether any of the risks are a determinant of trust, and/or the other way around. These relationships deserve further appropriately designed research.

Limitations

Our study has some limitations. First, this study was conducted during the COVID-19 pandemic, which complicated the data collection, as it became increasingly difficult to approach the targeted population. It became very difficult to recruit respondents who were unable or unwilling to fill out the web-based questionnaire. This may have resulted in a bias towards independently living older adults with better internet literacy. Second, we used a cross-sectional research design. Consequently, the causality of our findings cannot be claimed. Third, we noticed that some of the participants, especially those in the age group above 75 years, struggled to understand the use and utility of medical applications properly. To address this situation, the questionnaires and interviewers provided an additional explanation about medical applications. Finally, although the data are collected from a variety of contexts in the Netherlands, we cannot claim validity in other countries.

Recommendations and Future Research

The main contribution of this study is to provide the first large-scale quantitative evidence of the validity and significance of the perceived risk factors determining intention to use mHealth applications among older adults in the Netherlands. In our study, we identified trust and performance risk as the most important factors that had a relation with the older adults' intention to use mHealth services. We suggest involving older adults in the design and development of mHealth applications to ensure that the applications will be tailored to their needs and abilities. Moreover, we suggest involving medical specialists, geriatricians, and other experts in the development of these applications and making this explicit to potential users of the application to increase trust and diminish concerns about the performance of mHealth. Additionally, since privacy risks and legal concerns have a relation with the intention to use mHealth, we suggest that health care professionals, designers of mHealth applications, and the Dutch government use these findings to tailor their mHealth services and campaigns and address the concerns of older adults, to promote better adoption.

As we cannot confirm causality, we recommend studying perceived risk factors using controlled experiments rather than observational studies to confirm or disprove any potential causality of the relationships thus found. To further understand how the perceived risk factors explain behavioral intention to use, possibly via interaction with each other and the variables from models such as TAM, STAM, and UTAUT, qualitative studies are called for. Such qualitative studies can also enable effective solutions to eliminate barriers to medical application adoption. Furthermore, although the intention to use technology has been shown to predict actual usage [41], such experiments may research actual technology adoption, rather than the intention to use mHealth applications.

Acknowledgments

The authors would like to thank all the experts and older adults (participants) who helped us validate our questionnaire, as well as the participants and data assistants.

Authors' Contributions

MA designed the research project. NSK and MA developed the questionnaire. NSK, MA, and RvdB collected the data with the help of data assistants. NSK and RvdB performed the analyses under the supervision of MA. All the authors interpreted the results. RvdB and NSK wrote the initial version of the manuscript. All authors revised the paper critically.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Constructs and statements used in this study.
[DOCX File , 15 KB - mhealth_v9i8e26845_app1.docx]

Multimedia Appendix 2
Checklist for Reporting Results of Internet E-Surveys (CHERRIES).
[DOCX File , 23 KB - mhealth_v9i8e26845_app2.docx]

Multimedia Appendix 3
Results of linear regression analyses and correlation analyses.
[DOCX File , 26 KB - mhealth_v9i8e26845_app3.docx]

References

1. Prognosis population; gender and age, 2020-2060. StatLine. 2019 Dec 17. URL: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/84646NED/table?ts=1590564411512 [accessed 2020-07-13]
2. Life expectancy (observed) and prognosis 1950-2060. Public health care.info. 2019. URL: https://www.volksgezondheidenzorg.info/onderwerp/levensverwachting/cijfers-context/trends#!node-prognose-levensverwachting [accessed 2020-07-13]
3. World Population Ageing – Report 2017. New York: United Nations, Department of Economic and Social Affairs, Population Division; 2017. URL: https://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2017_Report.pdf [accessed 2019-10-06]
4. Nederlandse Zorgautoriteit. Monitor - Zorg voor ouderen 2018. Webpage in Dutch. 2018. URL: https://www.rijksoverheid.nl/documenten/rapporten/2018/04/19/monitor-zorg-voor-ouderen-2018 [accessed 2019-10-06]
5. Rynning E. The ageing populations of Europe--implications for health systems and patients' rights. Eur J Health Law 2008 Sep;15(3):297-306. [doi: 10.1163/157180908x338241] [Medline: 19024866]
6. Jeurissen P. Beheersing zorgkosten. Article in Dutch. Tijds. gezondheids.wetenschappen 2014 Apr 6;92(3):99-100. [doi: 10.1007/s12508-014-0039-6]
7. Global Observatory for eHealth. World Health Organization. 2017. URL: https://www.who.int/observatories/global-observatory-for-ehealth [accessed 2019-10-06]
8. Wouters M, Swinkels I, van Lettow B, de Jong J, Sinnige J, Brabers A. E-health in verschillende snelheden. Report in Dutch. eHealth-monitor 2018. Den Haag and Utrecht: Nictiz and het Nivel; 2018 Nov. URL: https://www.ehealth-monitor.nl/wp-content/themes/nictiz/assets/pdf/ehealth-monitor-2018.pdf [accessed 2019-10-06]
9. Sülz S, van Elten HJ, Askari M, Weggelaar-Jansen AM, Huijsman R. eHealth applications to support independent living of older persons: scoping review of costs and benefits identified in economic evaluations. J Med Internet Res 2021 Mar 09;23(3):e24363 [FREE Full text] [doi: 10.2196/24363] [Medline: 33687335]
10. Bruce CR, Harrison P, Nisar T, Giammattei C, Tan NM, Bliven C, et al. Assessing the impact of patient-facing mobile health technology on patient outcomes: retrospective observational cohort study. JMIR Mhealth Uhealth 2020 Jun 26;8(6):e19333 [FREE Full text] [doi: 10.2196/19333] [Medline: 32589161]
11. Salgado T, Tavares J, Oliveira T. Drivers of mobile health acceptance and use from the patient perspective: survey study and quantitative model development. JMIR Mhealth Uhealth 2020 Jul 09;8(7):e17588 [FREE Full text] [doi: 10.2196/17588] [Medline: 32673249]
12. Liu Y, Mao B, Liang S, Yang J, Lu H, Chai Y, Shanghai Clinical Treatment Experts Group for COVID-19. Association between age and clinical characteristics and outcomes of COVID-19. Eur Respir J 2020 May;55(5):2001112 [FREE Full text] [doi: 10.1183/13993003.01112-2020] [Medline: 32312864]
13. Dou K, Yu P, Deng N, Liu F, Guan Y, Li Z, et al. Patients' acceptance of smartphone health technology for chronic disease management: a theoretical model and empirical test. JMIR Mhealth Uhealth 2017 Dec 06;5(12):e177 [FREE Full text] [doi: 10.2196/mhealth.7886] [Medline: 29212629]
14. Breil B, Kremer L, Hennemann S, Apolinário-Hagen J. Acceptance of mHealth apps for self-management among people with hypertension. Stud Health Technol Inform 2019 Sep 03;267:282-288. [doi: 10.3233/SHTI190839] [Medline: 31483283]

15. Global diffusion of eHealth: making universal health coverage achievable. Report of the third global survey on eHealth. World Health Organisation. Geneva; 2016. URL: https://apps.who.int/iris/bitstream/handle/10665/252529/9789241511780-eng.pdf;jsessionid=C9C5ECC47AC7E602C18FDF7DE10AE136?sequence=1 [accessed 2020-07-22]
16. Atlas of eHealth country profiles: the use of eHealth in support of universal health coverage. WHO Global Observatory for eHealth. Geneva; 2016. URL: https://www.who.int/publications/i/item/9789241565219 [accessed 2019-01-08]
17. Adibi S. Mobile Health: A Technology Road Map. Switzerland: Springer International Publishing; 2015.
18. Chiarini G, Ray P, Akter S, Masella C, Ganz A. mHealth technologies for chronic diseases and elders: a systematic review. IEEE J Select Areas Commun 2013 Sep;31(9):6-18. [doi: 10.1109/jsac.2013.sup.0513001]
19. Helbostad J, Vereijken B, Becker C, Todd C, Taraldsen K, Pijnappels M, et al. Mobile health applications to promote active and healthy ageing. Sensors (Basel) 2017 Mar 18;17(3):622 [FREE Full text] [doi: 10.3390/s17030622] [Medline: 28335475]
20. Istepanian R, Woodward B. m-Health: Fundamentals and Applications. Hoboken, NJ, USA: John Wiley & Sons; 2016.
21. Malvey D, Slovensky DJ. mHealth: Transforming Healthcare. New York: Springer; 2014.
22. Changizi M, Kaveh MH. Effectiveness of the mHealth technology in improvement of healthy behaviors in an elderly population-a systematic review. Mhealth 2017;3:51 [FREE Full text] [doi: 10.21037/mhealth.2017.08.06] [Medline: 29430455]
23. Kampmeijer R, Pavlova M, Tambor M, Golinowska S, Groot W. The use of e-health and m-health tools in health promotion and primary prevention among older adults: a systematic literature review. BMC Health Serv Res 2016 Sep 05;16 Suppl 5:290 [FREE Full text] [doi: 10.1186/s12913-016-1522-3] [Medline: 27608677]
24. Fletcher J, Jensen R. Overcoming barriers to mobile health technology use in the aging population. Online J Nurs Inform 2015;19(3):1 [FREE Full text]
25. Spann A, Stewart E. Barriers and facilitators of older people's mHealth usage: a qualitative review of older people's views. Human Technology 2018 Nov 30:264-296. [doi: 10.17011/ht/urn.201811224834]
26. Agarwal R, Prasad J. Are individual differences germane to the acceptance of new information technologies? Decision Sciences 1999 Mar;30(2):361-391. [doi: 10.1111/j.1540-5915.1999.tb01614.x]
27. Harrison AW, Rainer RK. The influence of individual differences on skill in end-user computing. Journal of Management Information Systems 2015 Dec 16;9(1):93-111. [doi: 10.1080/07421222.1992.11517949]
28. Tams S, Grover V, Thatcher J. Modern information technology in an old workforce: toward a strategic research agenda. The Journal of Strategic Information Systems 2014 Dec;23(4):284-304. [doi: 10.1016/j.jsis.2014.10.001]
29. Askari M, Klaver NS, van Gestel TJ, van de Klundert J. Intention to use medical apps among older adults in the Netherlands: cross-sectional study. J Med Internet Res 2020 Sep 04;22(9):e18080 [FREE Full text] [doi: 10.2196/18080] [Medline: 32624465]
30. Conci M, Pianesi F, Zancanaro M. Useful, social and enjoyable: Mobile phone adoption by older people. INTERACT 2009:63 [FREE Full text]
31. Farage MA, Miller KW, Ajayi F, Hutchins D. Design principles to accommodate older adults. Glob J Health Sci 2012 Feb 29;4(2):2-25 [FREE Full text] [doi: 10.5539/gjhs.v4n2p2] [Medline: 22980147]
32. Mitzner TL, Boron JB, Fausset CB, Adams AE, Charness N, Czaja SJ, et al. Older adults talk technology: technology usage and attitudes. Comput Human Behav 2010 Nov 01;26(6):1710-1721 [FREE Full text] [doi: 10.1016/j.chb.2010.06.020] [Medline: 20967133]
33. Nikou S. Mobile technology and forgotten consumers: the young-elderly. Int J Consum Stud 2015 Apr 30;39(4):294-304. [doi: 10.1111/ijcs.12187]
34. Apolinário-Hagen J, Hennemann S, Fritsche L, Drüge M, Breil B. Determinant factors of public acceptance of stress management apps: survey study. JMIR Ment Health 2019 Nov 07;6(11):e15373 [FREE Full text] [doi: 10.2196/15373] [Medline: 31697243]
35. 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]
36. Zhang Y, Liu C, Luo S, Xie Y, Liu F, Li X, et al. Factors influencing patients' intentions to use diabetes management apps based on an extended unified theory of acceptance and use of technology model: web-based survey. J Med Internet Res 2019 Aug 13;21(8):e15023 [FREE Full text] [doi: 10.2196/15023] [Medline: 31411146]
37. Alaiad A, Zhou L. The determinants of home healthcare robots adoption: an empirical investigation. Int J Med Inform 2014 Nov;83(11):825-840. [doi: 10.1016/j.ijmedinf.2014.07.003] [Medline: 25132284]
38. Kim DJ, Ferrin DL, Rao HR. A trust-based consumer decision-making model in electronic commerce: the role of trust, perceived risk, and their antecedents. Decision Support Systems 2008 Jan;44(2):544-564. [doi: 10.1016/j.dss.2007.07.001]
39. Guo X, Han X, Zhang X, Dang Y, Chen C. Investigating m-Health acceptance from a protection motivation theory perspective: gender and age differences. Telemed J E Health 2015 Aug;21(8):661-669. [doi: 10.1089/tmj.2014.0166] [Medline: 25919800]
40. Cajita MI, Hodgson NA, Budhathoki C, Han H. Intention to use mHealth in older adults with heart failure. J Cardiovasc Nurs 2017;32(6):E1-E7. [doi: 10.1097/jcn.0000000000000401]

41. Holden RJ, Karsh B. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
42. 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]
43. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterly 2003 Nov;27(3):425-478. [doi: 10.2307/30036540]
44. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Management Science 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
45. Fischer SH, David D, Crotty BH, Dierks M, Safran C. Acceptance and use of health information technology by community-dwelling elders. Int J Med Inform 2014 Sep;83(9):624-635 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.06.005] [Medline: 24996581]
46. Rasche P, Wille M, Bröhl C, Theis S, Schäfer K, Knobe M, et al. Prevalence of health app use among older adults in Germany: national survey. JMIR Mhealth Uhealth 2018 Jan 23;6(1):e26 [FREE Full text] [doi: 10.2196/mhealth.8619] [Medline: 29362211]
47. Parker SJ, Jessel S, Richardson JE, Reid MC. Older adults are mobile too! Identifying the barriers and facilitators to older adults' use of mHealth for pain management. BMC Geriatr 2013 May 06;13:43 [FREE Full text] [doi: 10.1186/1471-2318-13-43] [Medline: 23647949]
48. Ferguson SJ, Goodwin AD. Optimism and well-being in older adults: the mediating role of social support and perceived control. Int J Aging Hum Dev 2010 Jun 21;71(1):43-68. [doi: 10.2190/ag.71.1.c]
49. Vaportzis E, Clausen MG, Gow AJ. Older adults perceptions of technology and barriers to interacting with tablet computers: a focus group study. Front Psychol 2017 Oct 04;8:1687 [FREE Full text] [doi: 10.3389/fpsyg.2017.01687] [Medline: 29071004]
50. Henry TA. What do mHealth apps share about users? Guidelines demand clarity. American Medical Association. 2019 Dec 19. URL: https://www.ama-assn.org/practice-management/digital/what-do-mhealth-apps-share-about-users-guidelines-demand-clarity [accessed 2020-06-16]
51. Guo X, Zhang X, Sun Y. The privacy–personalization paradox in mHealth services acceptance of different age groups. Electronic Commerce Research and Applications 2016 Mar;16:55-65. [doi: 10.1016/j.elerap.2015.11.001]
52. Zhao Y, Ni Q, Zhou R. What factors influence the mobile health service adoption? A meta-analysis and the moderating role of age. International Journal of Information Management 2018;43:342-350.
53. Deng Z, Liu S, Hinz O. The health information seeking and usage behavior intention of Chinese consumers through mobile phones. Information Technology & People 2015;28:405-423. [doi: 10.1108/ITP-03-2014-0053]
54. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
55. Venkatesh V. Determinants of perceived ease of use: integrating control, intrinsic motivation, and emotion into the technology acceptance model. Information Systems Research 2000 Dec;11(4):342-365. [doi: 10.1287/isre.11.4.342.11872]
56. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ 2011 Jun 27;2:53-55 [FREE Full text] [doi: 10.5116/ijme.4dfb.8dfd] [Medline: 28029643]
57. Cortina JM. What is coefficient alpha? an examination of theory and applications. Journal of Applied Psychology 1993;78(1):98-104. [doi: 10.1037/0021-9010.78.1.98]
58. Pallant J. SPSS Survival Manual. A step by step guide to data analysis using SPSS. 4th edition. Crows Nest NSW: Allen & Unwin; 2011.
59. Chen K, Chan AHS. Gerontechnology acceptance by elderly Hong Kong Chinese: a senior technology acceptance model (STAM). Ergonomics 2014;57(5):635-652. [doi: 10.1080/00140139.2014.895855] [Medline: 24655221]
60. Wildenbos G, Peute L, Jaspers M. Aging barriers influencing mobile health usability for older adults: a literature based framework (MOLD-US). Int J Med Inform 2018 Jun;114:66-75. [doi: 10.1016/j.ijmedinf.2018.03.012] [Medline: 29673606]
61. Nunes A, Limpo T, Castro SL. Acceptance of mobile health applications: examining key determinants and moderators. Front Psychol 2019;10:2791 [FREE Full text] [doi: 10.3389/fpsyg.2019.02791] [Medline: 31920836]
62. Gustafson PE. Gender differences in risk perception: theoretical and methodological perspectives. Risk Anal 1998 Dec;18(6):805-811. [doi: 10.1023/b:rian.0000005926.03250.c0] [Medline: 9972583]
63. Li J, Ma Q, Chan AH, Man S. 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]
64. Martin N. The Ego Unmasked: Meeting the Greatest Challenge of Your Life. Pittsburgh: Dorrance Publishing Co; 2010.
65. Verhagen L. How effective is the corona app? And what about privacy? de Volkskrant. 2020 Jul 16. URL: https://tinyurl.com/jya6synm [accessed 2020-10-06]
66. Onderzoeksrapportage bron- en contactopsporingsapps. Article in Dutch. Autoriteit Persoonsgegevens. 2020 Apr 20. URL: https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/onderzoeksrapportage_bron-_en_contactopsporingsapps.pdf [accessed 2020-10-06]
67. Siegrist M, Gutscher H, Earle TC. Perception of risk: the influence of general trust, and general confidence. Journal of Risk Research 2006 Aug 15;8(2):145-156. [doi: 10.1080/1366987032000105315]

68. Rolison J, Hanoch Y, Freund A. Perception of risk for older adults: differences in evaluations for self versus others and across risk domains. Gerontology 2019;65(5):547-559. [doi: 10.1159/000494352] [Medline: 30566940]
69. Jonker M, de Bekker-Grob E, Veldwijk J, Goossens L, Bour S, Rutten-Van Mölken M. COVID-19 contact tracing apps: predicted uptake in the Netherlands based on a discrete choice experiment. JMIR Mhealth Uhealth 2020 Oct 09;8(10):e20741 [FREE Full text] [doi: 10.2196/20741] [Medline: 32795998]
70. Schnall R, Higgins T, Brown W, Carballo-Dieguez A, Bakken S. Trust, perceived risk, perceived ease of use and perceived usefulness as factors related to mHealth technology use. Stud Health Technol Inform 2015;216:467-471 [FREE Full text] [Medline: 26262094]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
mHealth: mobile health
TAM: technology acceptance model
UTAUT: unified theory of acceptance and use of technology

Edited by L Buis; submitted 30.12.20; peer-reviewed by B Breil, N Muinga; comments to author 15.01.21; revised version received 11.03.21; accepted 30.03.21; published 30.08.21.

Please cite as:
Klaver NS, van de Klundert J, van den Broek RJGM, Askari M
Relationship Between Perceived Risks of Using mHealth Applications and the Intention to Use Them Among Older Adults in the Netherlands: Cross-sectional Study
JMIR Mhealth Uhealth 2021;9(8):e26845
URL: https://mhealth.jmir.org/2021/8/e26845
doi:10.2196/26845
PMID:34459745

Viewpoint

Mobile Apps as Audience-Centered Health Communication Platforms

Michael Mackert[1,2], PhD; Dorothy Mandell[3,4], PhD; Erin Donovan[5], PhD; Lorraine Walker[6], EdD, MPH; Mike Henson-García[2,7], MPH; Lindsay Bouchacourt[1], MA

[1]Stan Richards School of Advertising and Public Relations, The University of Texas at Austin, Austin, TX, United States
[2]Center for Health Communication, The University of Texas at Austin, Austin, TX, United States
[3]School of Community and Rural Health, University of Texas Health Science Center at Tyler, Tyler, TX, United States
[4]Population Health, The University of Texas System, Austin, TX, United States
[5]Department of Communication Studies, The University of Texas at Austin, Austin, TX, United States
[6]School of Nursing, The University of Texas at Austin, Austin, TX, United States
[7]School of Public Health, University of Texas Health Science Center - Dallas Regional Campus, Dallas, TX, United States

Corresponding Author:
Michael Mackert, PhD
Stan Richards School of Advertising and Public Relations
The University of Texas at Austin
300 W Dean Keeton St
Austin, TX, 78712
United States
Phone: 1 512 348 8490
Email: mackert@utexas.edu

Abstract

Health communication campaigns often suffer from the shortcomings of a limited budget and limited reach, resulting in a limited impact. This paper suggests a shift of these campaigns to audience-centered communication platforms—particularly, apps on mobile phones. By using a common platform, multiple interventions and campaigns can combine resources and increase user engagement, resulting in a larger impact on health behavior. Given the widespread use of mobile phones, mobile apps can be an effective and efficient tool to provide health interventions. One such platform is Father's Playbook, a mobile app designed to encourage men to be more involved during their partner's pregnancy. Health campaigns and interventions looking to reach expectant fathers can use Father's Playbook as a vehicle for their messages.

(JMIR Mhealth Uhealth 2021;9(8):e25425) doi:10.2196/25425

KEYWORDS

health communication; mHealth; mobile apps; mobile health; prenatal health; pregnancy; audience-centered

Introduction

There are numerous public health issues, ranging from disparities in maternal mortality to the reframing of child abuse as a population health concern, where communication plays a key part in the solution. For example, the US Centers for Disease Control and Prevention's *Tips From Former Smokers* was a national mass media anti-smoking campaign that profiled real people living with serious long-term health effects from smoking [1]. The campaign is known for its television spots with graphic and emotional testimonials from former smokers. This massive campaign helped approximately 1 million people successfully quit smoking [1]. A health communication example using new media is the Text4Baby smartphone app, which is aimed toward expectant and new mothers. The app provides information on topics ranging from baby milestones to nutrition to childcare tips, and it sends over 250 SMS text messages to the user's phone with the most critical information for pregnant women and mothers [2]. Women who used the Text4Baby app felt more prepared to be new mothers [3], had higher attitudes toward prenatal vitamins [4], and had a higher level of pregnancy health knowledge [4].

Although high-profile and effective health communication interventions exist, it must be acknowledged that many do not achieve their objectives. A primary reason could be that health communication interventions are "underdosed"—most simply do not have the budget to reach enough people to make a

substantial difference. Although the *Tips From Former Smokers* campaign was very successful, it had a budget of US $48 million in 2012 [1], which is anomalously high among health campaign budgets. This applies to health communication campaigns that focus on broad message dissemination as well as to more complex and interactive interventions.

Health communication could benefit from a fundamental paradigm shift. Interventions are often designed appropriately by following best practices [5,6], working within the budget and evaluation constraints of a particular project. To improve the efficacy of health communication interventions, we suggest a shift to audience-centered communication platforms, which are platforms that focus on a specific audience but tailor their content to address different subject matter (in this case, health issues) or different subsets of the audience.

When considering the potential of an audience-centered platform, mobile apps specifically can be extremely effective due to their versatility in content and widespread reach. Mobile apps could provide opportunities for multiple interventions and campaigns to combine resources and increase user engagement in service of promoting health behavior change and public health. As just one example, Bright by Text, a smartphone app for parents that provides information about early childhood topics, has regional implementations [7]. These regional implementations provide location-specific activities for parents to do with their children; this geographic tailoring can make the content more specific and useful for parents who live in different regions across the United States and avoids the need for regional programs to find their own ways to spread information to those parents.

Given the widespread use of mobile devices in the general public [8], delivering health messages through mobile apps is particularly useful [9]. Mobile apps can efficiently deliver appropriate doses of health messages along with providing "in-the-moment" health interventions [10,11], which are essentially a type of just-in-time adaptive intervention [12]. Mobile apps can deliver tailored health messages to people in specific and opportune moments within their everyday settings, effectively addressing the "message dosing issues" many health campaigns face [10,11]. Mobile apps can also effectively intervene among multiple health issues in one specific group or population. Health communication interventions are often built around a single health issue, despite the fact that many health issues coincide together within specific groups of individuals [13-15]. The time and resources used on promoting many independent health communication interventions could be better spent if there was a common platform that could be shared, especially if the interventions were all targeting a common audience.

One such platform to share health information is Father's Playbook, a smartphone app designed for men to use during and after their partner's pregnancy. Health communication interventions and campaigns attempting to reach expectant fathers can use Father's Playbook as a way to reach an audience that already exists.

Father's Playbook Case Study

The Father's Playbook smartphone app was created to fill a large gap in pregnancy-related health information—the lack of male-focused pregnancy information and resources. While women should be the main focus during pregnancy, including men has benefits for father, mother, and baby. Father involvement during the prenatal stage has shown to have many positive outcomes, including increased communication between partners, higher number of provider visits during pregnancy, and increased postpartum best practices [16,17]. Although men do feel strongly about being involved in pregnancy health [18], structural health care barriers and personal barriers may prevent them from being engaged. To help combat these barriers, Father's Playbook was designed to engage men in prenatal health.

The Father's Playbook app was developed using an incremental approach to the research and the app's development. This process was cost-effective, and it allowed us to learn from mistakes and build on previous knowledge. Before the Father's Playbook smartphone app existed, a web-based pilot was created. To obtain more generalizable results regarding attitudes and barriers to prenatal health involvement, a survey with a nationally representative sample of men was conducted [18]. Overall, men believe that it is important to be involved in pregnancy health; however, perceived barriers (eg, time restraints, unclear role, financial burdens) still exist [18]. This survey also required participants to interact with the website, and the participants made suggestions on how to improve the site. Through interviews and the survey, we were able to obtain men's opinions on the website, along with suggestions for features that would improve the website [19]. This information allowed us to prioritize features to be developed (articles and an interactive budget calculator) and focus on user testing for new features.

Both the app's development and future research related to the app have next steps. Presently, the app's content is available in English and Spanish, which opens the door to a larger audience of men. Following this type of audience specification, in the future, the goal is to tailor content toward different types of fathers. For example, the experience of a stay-at-home father versus a single parent father greatly differs, and the pregnancy and fatherhood experience of transgender men who become pregnant will differ from the experience of a cisgender man [20]. On the research side, the next step will be to focus on the father's engagement in prenatal health in a broader sense. To be able to measure the effectiveness of Father's Playbook in improving father engagement, baseline data is needed on the current level of father engagement in prenatal health. As such, a representative survey targeting fathers and expectant fathers is being conducted to gather baseline data. Future development of the app can strengthen its potential as a platform to enable more efficient communication with expectant fathers rather than individual programs and efforts to reach this audience.

There are a wide variety of opportunities to improve paternal, maternal, and child health through improved communication with expectant fathers. There is a wealth of evidence showing

that fathers want to be more engaged in the pregnancy and feel underprepared for fatherhood [19,21,22]. However, there is a dearth of research focused on how to engage and prepare fathers. The app provides a flexible platform to test education and communication strategies with fathers directly. For instance, directed education about nutrition and pregnancy complications may result in men assisting their pregnant partners in making healthier nutrition choices and may increase their ability to identify obstetric emergencies. Evidence also suggests that when expectant parents are unsatisfied with their partnerships, they are more likely to exhibit insensitive parenting styles once their infant is born. The app may provide a platform for disseminating communication interventions to help the quality of the relationship between expecting couples [23]. Additionally, studies have found that low levels of paternal engagement throughout the prenatal period are related to birth complications [24] and to low postnatal engagement. Positive father involvement in early childhood is tied to a variety of positive cognitive and health outcomes for children [25]; therefore, increasing the father's engagement has the potential to result in positive health outcomes across the entire family unit.

Father's Playbook as an Audience-Centered Platform

Father's Playbook was designed using audience-centered principles, which means that the content, language, and approach used in the app are tailored to best fit the audience (new and expectant fathers). By focusing on the audience, the app can tailor information and its delivery to the needs of the audience. In its current format, Father's Playbook is a single app that includes content pages about specific father-related information and interactive app features (such as a budget calculator) to encourage expectant fathers to become more engaged during and after their partner's pregnancy. Future versions of the app will include tailored content, allowing the app to be more personally relevant to its user. This ability to tailor content can allow other health communication scholars to use the app to deploy tailored interventions and campaigns to the target audience of expectant and new fathers.

The shift to thinking of audience-centered platforms can broaden the reach and efficiency of interventions that have successfully developed an approach for reaching a particular audience. As an example, although Father's Playbook is currently focused on amplifying the engagement of expectant fathers during the prenatal period, there are other issues that commonly affect expectant fathers, and at any time, the app can deploy other types of content to address the audience's other public health needs. Other examples of public health information expectant fathers might want to consume include management of nutrition and physical activity throughout pregnancy, prevention of paternal postpartum depression, smoking cessation, and information related to paternity testing. These different health issues can be addressed specifically through tailored content in the app, without the need for researchers and professionals with interests in these issues to develop an entirely new intervention or campaign and then determine how to reach the target audience.

Along with addressing health issues of expectant and new fathers, Father's Playbook can be used to address parenting issues and child-rearing best practices. For example, a researcher could be interested in testing an intervention that increases the amount of time a father reads to his children; as part of the intervention, participants would need to download the Father's Playbook app. The app could add features that specifically work to reach the intervention objectives, such as content articles that discuss the benefits of fathers reading to their children, interactive games that encourage book reading, and a list of book suggestions appropriate for specific age ranges.

Father's Playbook was built with the goal of increasing father engagement during pregnancy. Now that an audience of expectant and new fathers exists, other researchers and practitioners can access this unique audience to address a myriad of parenting and health issues, allowing a collaborative approach toward health communication campaigns and interventions. Without the use of an audience-centered platform, health professionals would only need to use their own resources to reach this audience.

Conclusion

Health communication can be an effective tool to help improve the health and well-being of individuals and populations. There is a strong evidence base of health communication that can be leveraged across health issues and audiences, such as the increased efficacy of targeted and tailored messages compared to more general appeals [26].

Despite the high-profile success of health communication campaigns that have achieved important and demonstrated benefits, substantial opportunities remain to advance the field through new approaches to message design and reaching audiences. The approach advocated in this paper is to focus on more audience-focused platforms, which could be a more efficient strategy for message dissemination. Taking an audience-centered approach allows for a better understanding of people, their behaviors, their contexts, and their intersections allowing for more nuanced health communication and health promotion efforts. Given that disease manifests from the compound of multiple risk factors—and said factors are differentially distributed across various lifestyles and identities—audience-centered approaches have the potential to be highly effective vehicles for health transformation and useful for any number of audiences, ranging from transgender women to Black men who have sex with men to older persons managing multiple chronic conditions to COVID-19 survivors.

It is a well-established principle of health communication that targeted and tailored communication is more effective than general messages. The approach advocated in this paper—to build audience-centered communication platforms—is a promising approach to develop more cost-effective, engaging, and effective health communication interventions.

Acknowledgments

This article is based on work resulting from participation in the Puentes 2.0 Conference sponsored by the Department of Population Health, the Lozano Long Institute of Latin-American Studies (LLILAS)–Benson Mexico Center, the Office of the Vice President of Research, and University of Texas System Collaboration for Population Health Innovation and Improvement (CoPHII) project. This project was funded through the Texas Safe Babies project (IAC 24307170) from the Texas Department of Family and Protective Services. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Texas Department of Family and Protective Services.

Conflicts of Interest

None declared.

References

1. Tips From Former Smokers. US Centers for Disease Control and Prevention. URL: https://www.cdc.gov/tips [accessed 2020-10-21]
2. Text4Baby. URL: https://www.text4baby.org [accessed 2020-10-21]
3. Evans WD, Wallace JL, Snider J. Pilot evaluation of the text4baby mobile health program. BMC Public Health 2012 Nov 26;12:1031 [FREE Full text] [doi: 10.1186/1471-2458-12-1031] [Medline: 23181985]
4. Evans WD, Wallace Bihm J, Szekely D, Nielsen P, Murray E, Abroms L, et al. Initial outcomes from a 4-week follow-up study of the Text4baby program in the military women's population: randomized controlled trial. J Med Internet Res 2014 May 20;16(5):e131 [FREE Full text] [doi: 10.2196/jmir.3297] [Medline: 24846909]
5. Making Health Communication Programs Work (Pink Book). National Cancer Institute. URL: https://www.cancer.gov/publications/health-communication/pink-book.pdf [accessed 2020-10-21]
6. Mackert M, Lazard A, Love B. Designing Effective Health Messages. Dubuque, IA: Kendall Hunt; 2017.
7. Bright by Text. URL: https://brightbytext.org [accessed 2021-04-27]
8. Mobile Fact Sheet: Internet/Broadband, Mobile, Social Media. Pew Research Center. 2019. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2020-10-21]
9. Riley W, Rivera D, Atienza A, Nilsen W, Allison S, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
10. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010;15(1):1-39 [FREE Full text] [doi: 10.1348/135910709x466063]
11. Runyan J, Steinke EG. Virtues, ecological momentary assessment/intervention and smartphone technology. Front Psychol 2015;6:481 [FREE Full text] [doi: 10.3389/fpsyg.2015.00481] [Medline: 25999869]
12. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
13. Howell EA. Reducing disparities in severe maternal morbidity and mortality. Clin Obstet Gynecol 2018 Jun;61(2):387-399 [FREE Full text] [doi: 10.1097/GRF.0000000000000349] [Medline: 29346121]
14. Pérez C, Sánchez H, Ortiz A. Prevalence of overweight and obesity and their cardiometabolic comorbidities in Hispanic adults living in Puerto Rico. J Community Health 2013 Dec;38(6):1140-1146 [FREE Full text] [doi: 10.1007/s10900-013-9726-5] [Medline: 23846388]
15. Sanna L, Stuart A, Pasco J, Kotowicz M, Berk M, Girardi P, et al. Physical comorbidities in men with mood and anxiety disorders: a population-based study. BMC Med 2013 Apr 24;11:110 [FREE Full text] [doi: 10.1186/1741-7015-11-110] [Medline: 23618390]
16. Zvara B, Schoppe-Sullivan S, Dush C. Fathers' involvement in child health care: associations with prenatal involvement, parents' beliefs, and maternal gatekeeping. Fam Relat 2013 Oct;62(4):649-661 [FREE Full text] [doi: 10.1111/fare.12023] [Medline: 26405366]
17. Mullany B, Becker S, Hindin M. The impact of including husbands in antenatal health education services on maternal health practices in urban Nepal: results from a randomized controlled trial. Health Educ Res 2007 Apr;22(2):166-176 [FREE Full text] [doi: 10.1093/her/cyl060] [Medline: 16855015]
18. Mackert M, Guadagno M, Lazard A, Donovan E, Rochlen A, Garcia A, et al. Engaging men in prenatal health via eHealth: findings from a national survey. JMIR Pediatr Parent 2018 Aug 09;1(2):1-7 [FREE Full text] [doi: 10.2196/pediatrics.9513] [Medline: 31518311]
19. Mackert M, Guadagno M, Lazard A, Donovan E, Rochlen A, Garcia A, et al. Engaging men in prenatal health promotion: a pilot evaluation of targeted e-health content. Am J Mens Health 2017 May;11(3):719-725 [FREE Full text] [doi: 10.1177/1557988316679562] [Medline: 27956587]
20. Light A, Obedin-Maliver J, Sevelius J, Kerns J. Transgender men who experienced pregnancy after female-to-male gender transitioning. Obstet Gynecol 2014 Dec;124(6):1120-1127. [doi: 10.1097/AOG.0000000000000540] [Medline: 25415163]

21. Kaye D, Kakaire O, Nakimuli A, Osinde M, Mbalinda S, Kakande N. Male involvement during pregnancy and childbirth: men's perceptions, practices and experiences during the care for women who developed childbirth complications in Mulago Hospital, Uganda. BMC Pregnancy Childbirth 2014 Jan 31;14:54 [FREE Full text] [doi: 10.1186/1471-2393-14-54] [Medline: 24479421]
22. Deave T, Johnson D. The transition to parenthood: what does it mean for fathers? J Adv Nurs 2008 Sep;63(6):626-633 [FREE Full text] [doi: 10.1111/j.1365-2648.2008.04748.x] [Medline: 18808584]
23. Cabrera N, Shannon J, Mitchell S, West J. Mexican American mothers and fathers' prenatal attitudes and father prenatal involvement: links to mother-infant interaction and father engagement. Sex Roles 2009 Apr 01;60(7-8):510-526 [FREE Full text] [doi: 10.1007/s11199-008-9576-2] [Medline: 20300483]
24. Lee S, Sanchez D, Grogan-Kaylor A, Lee J, Albuja A. Father early engagement behaviors and infant low birth weight. Matern Child Health J 2018 Oct;22(10):1407-1417 [FREE Full text] [doi: 10.1007/s10995-018-2521-2] [Medline: 29564605]
25. Cabrera N, Volling B, Barr R. Fathers are parents, too! Widening the lens on parenting for children's development. Child Dev Perspect 2018 Jan 08;12(3):152-157 [FREE Full text] [doi: 10.1111/cdep.12275]
26. Schmid K, Rivers S, Latimer A, Salovey P. Targeting or tailoring? Maximizing resources to create effective health communications. Mark Health Serv 2008;28(1):32-37 [FREE Full text] [Medline: 18389854]

Abbreviations

CoPHII: Collaboration for Population Health Innovation and Improvement
LLILAS: Lozano Long Institute of Latin-American Studies

Edited by L Buis; submitted 01.11.20; peer-reviewed by T Mirzaei, AV Das; comments to author 11.03.21; revised version received 06.05.21; accepted 23.07.21; published 17.08.21.

Please cite as:
Mackert M, Mandell D, Donovan E, Walker L, Henson-García M, Bouchacourt L
Mobile Apps as Audience-Centered Health Communication Platforms
JMIR Mhealth Uhealth 2021;9(8):e25425
URL: https://mhealth.jmir.org/2021/8/e25425
doi:10.2196/25425
PMID:34402797

Original Paper

Effects of a Personalized Smartphone App on Bowel Preparation Quality: Randomized Controlled Trial

Quirine E W van der Zander[1,2], MSc, MD; Ankie Reumkens[1,3], MSc, MD; Bas van de Valk[1], MSc, MD; Bjorn Winkens[4], PhD; Ad A M Masclee[1,3], MD, Prof Dr; Rogier J J de Ridder[1], MSc, MD

[1]Division of Gastroenterology and Hepatology, Department of Internal Medicine, Maastricht University Medical Center, Maastricht, Netherlands
[2]GROW, School for Oncology and Developmental Biology, Maastricht University, Maastricht, Netherlands
[3]NUTRIM, School of Nutrition and Translational Research in Metabolism, Maastricht University, Maastricht, Netherlands
[4]Department of Methodology and Statistics, Maastricht University, Maastricht, Netherlands

Corresponding Author:
Quirine E W van der Zander, MSc, MD
GROW, School for Oncology and Developmental Biology
Maastricht University
Universiteitssingel 40
Maastricht, 6229 ER
Netherlands
Phone: 31 433875021
Email: q.vanderzander@maastrichtuniversity.nl

Abstract

Background: Adequate bowel preparation is essential for the visualization of the colonic mucosa during colonoscopy. However, the rate of inadequate bowel preparation is still high, ranging from 18% to 35%; this may lead to a higher risk of missing clinically relevant lesions, procedural difficulties, prolonged procedural time, an increased number of interval colorectal carcinomas, and additional health care costs.

Objective: The aims of this study are to compare bowel preparation instructions provided via a personalized smartphone app (Prepit, Ferring B V) with regular written instructions for bowel preparation to improve bowel preparation quality and to evaluate patient satisfaction with the bowel preparation procedure.

Methods: Eligible patients scheduled for an outpatient colonoscopy were randomized to a smartphone app group or a control group. Both the groups received identical face-to-face education from a research physician, including instructions about the colonoscopy procedure, diet restrictions, and laxative intake. In addition, the control group received written information, whereas the smartphone app group was instructed to use the smartphone app instead of the written information for the actual steps of the bowel preparation schedule. All patients used bisacodyl and sodium picosulfate with magnesium citrate as laxatives. The quality of bowel preparation was scored using the Boston Bowel Preparation Scale (BBPS) by blinded endoscopists. Patient satisfaction was measured using the Patient Satisfaction Questionnaire-18.

Results: A total of 87 patients were included in the smartphone app group and 86 in the control group. The mean total BBPS score was significantly higher in the smartphone app group (mean 8.3, SD 0.9) than in the control group (mean 7.9, SD 1.2; *P*=.03). The right colon showed a significantly higher bowel preparation score in the smartphone app group (mean 2.7, SD 0.5 vs mean 2.5, SD 0.6; *P*=.04). No significant differences were observed in segment scores for the mean transverse colon (mean 2.8, SD 0.4 vs mean 2.8, SD 0.4; *P*=.34) and left colon (mean 2.8, SD 0.4 vs mean 2.6, SD 0.5; *P*=.07). General patient satisfaction was high for the smartphone app group (mean 4.4, SD 0.7) but showed no significant difference when compared with the control group (mean 4.3, SD 0.8; *P*=.32).

Conclusions: Our personalized smartphone app significantly improved bowel preparation quality compared with regular written instructions for bowel preparation. In particular, in the right colon, the BBPS score improved, which is of clinical relevance because the right colon is considered more difficult to clean and the polyp detection rate in the right colon improves with improvement of bowel cleansing of the right colon. No further improvement in patient satisfaction was observed compared with patients receiving regular written instructions.

Trial Registration: ClinicalTrials.gov NCT03677050; https://clinicaltrials.gov/ct2/show/NCT03677050

(JMIR Mhealth Uhealth 2021;9(8):e26703) doi:10.2196/26703

KEYWORDS

colonoscopy; laxatives; bowel preparation; smartphone application; smartphone; patient satisfaction; randomized controlled trial; mobile phone; mHealth

Introduction

Background

Colonoscopy is considered the gold standard for diagnosing colorectal pathologies. The efficacy and safety of colonoscopy are related to the quality of the preinvestigational bowel preparation. Adequate bowel preparation is essential for the optimal visualization of the colonic mucosa during colonoscopy. Inadequate bowel preparation is associated with the risk of missing clinically relevant lesions, procedural difficulties, prolonged procedural time, an increased number of interval colorectal carcinomas, and additional health care costs [1-6]. Currently reported rates of inadequate bowel preparation range from 18% to 35% [1,7], leaving room for improvement.

Previous studies have evaluated various factors that can negatively affect bowel preparation, such as dietary restrictions (low-fiber vs clear liquid diet), laxative administration (single vs split dose), inadequate information precolonoscopy, and long waiting times [8-12]. In addition, bowel preparation quality depends on patients' tolerability to the laxative and patients' satisfaction. Patient satisfaction is inherently correlated with patients' compliance with the physician-recommended bowel preparation schedules.

Strategies to improve bowel preparation aim to inform patients more extensively about the preparation procedure and remind patients when action is needed (ie, start of diet modifications and intake of the laxative). Several of these strategies, including visual aids, educational videos, and SMS reminders, have provided better bowel preparation quality when compared with regular instructions [13]. Current colonoscopy preparation guidelines recommend providing patients with both verbal and written instructions and acknowledge the added value of providing educational booklets [14,15].

Objectives

A new method for informing and instructing patients is via a personalized smartphone app. In 2017, 93% of Dutch adults possessed a smartphone. The highest percentage of smartphone use was found in the younger age groups, but 90% of people aged ≥55 years had access to a smartphone [16]. Therefore, this technology has the potential to improve bowel preparation quality during colonoscopy. This study aims to investigate the quality of bowel preparation and patient satisfaction in patients using a newly developed, personalized smartphone app in addition to verbal instructions compared with regular verbal and written instructions.

Methods

Study Design

This prospective, endoscopist-blinded, randomized controlled trial was conducted at the Maastricht University Medical Center+, Maastricht, the Netherlands, from August 2018 to November 2019. The study was conducted in accordance with the Declaration of Helsinki [17] and the General Data Protection Regulation [18]. The Medical Ethical Review Committee of the Maastricht University Medical Center (MEC 16-4-141) approved the study. This study is registered at ClinicalTrials.gov (NCT03677050).

Subjects

Patients who were aged ≥18 years, who possessed a smartphone, who were referred to the outpatient clinic for a colonoscopy screening visit by their general practitioner or by the Dutch colorectal cancer screening program, and who were prescribed sodium picosulfate with magnesium citrate (SPMC) were eligible to participate. Hospitalized patients, patients undergoing an emergency colonoscopy, and patients without a smartphone were not considered eligible for participation. All patients fulfilling these inclusion and exclusion criteria were considered for inclusion in this study, and all included patients provided written informed consent. No incentives were offered to participating patients.

Randomization and Group Description

Patient education occurred during a screening visit at the outpatient clinic 1-4 weeks before colonoscopy. During this visit, patients were randomly assigned to the smartphone app group or the control group using a computer-generated randomization list in a 1:1 sequence based on the order of inclusion. Patients from both the groups received a hyperlink to a web-based educational video explaining the colonoscopy procedure. Patients in the control group received verbal and written information concerning diet restrictions, bowel preparation schedules, and laxatives. Patients in the smartphone app group had to install the app on their Android or iOS smartphones, which was accessible by a quick response code (Prepit, Ferring B V; for the Consolidated Standards of Reporting Trials, see Multimedia Appendix 1 [1,7,16,19-25]). Instead of written instructions, patients in the smartphone app group received information and instructions via the smartphone app. The information and instructions provided via the smartphone app were similar to the written instructions of the control group. However, the information was presented in a more visual way, that is, providing pictograms of low-fiber food products and images of the desired stool consistency after ingestion of the laxatives. Furthermore, the smartphone app provided the patients with personalized notifications about the steps of bowel preparation tailored to the exact colonoscopy date and time (Figure 1). It did not take extra time to provide the explanation via the smartphone app compared with the explanation given via the written instructions. Patient satisfaction with the bowel preparation procedure was evaluated using a self-assessed paper questionnaire, the Patient Satisfaction Questionnaire-18 (PSQ-18). This questionnaire was handed out

to the patients during the screening visit and filled in by the patients on the day of the colonoscopy. Patients completed the questionnaire before the colonoscopy was performed, as the actual experience of undergoing the colonoscopy was not asked for and could possibly (both negatively and positively) influence patient satisfaction regarding the bowel preparation procedure (for the questionnaire, please refer to Multimedia Appendix 2 [26]).

Figure 1. Smartphone app screenshots. (A) date and time entry, (B) educational tools, (C) date and time specific bowel preparation schedule, (D) examples of low-fiber diet, (E) picoprep preparation instructions, and (F) examples of clear liquids. Copyright Prepit, Ferring B V.

Bowel Preparation Schedule and Instructions

Instructions were delivered face-to-face by 2 research physicians (QEWVDZ and BVDV). Patients were instructed to follow a low-fiber diet 2 days before the colonoscopy. All patients were prescribed SPMC in a split-dose regimen of 2 doses, consisting of 10.0 mg sodium picosulfate, 3.5 g magnesium oxide, and 12.0 g citric acid (Picoprep, Ferring B V). Patients scheduled for a colonoscopy in the morning or early afternoon were instructed to take the first SPMC dose the evening before and

the second dose the morning of the colonoscopy. For colonoscopies scheduled in the afternoon, patients had to take both SPMC doses the morning of the examination, with a 2- to 5-hour interval between both the doses. All patients were also administered 10.0 mg of bisacodyl as an additive to the first SPMC dose.

Outcomes

The primary outcome was bowel preparation quality assessed using the Boston Bowel Preparation Scale (BBPS). The BBPS is a validated and reliable scale that rates bowel cleanliness for each colonic segment (right, transverse, and left) after washing, suctioning, and cleaning maneuvers have been performed by the endoscopist [27]. Each segment is scored on a scale from 0 to 3 (3 being the cleanest) [28,29]. Segment scores were summed to calculate the total BBPS, which ranged from 0 to 9. Bowel preparation was considered adequate when the total score was ≥6 and all segment scores were ≥2. This cut-off value has been shown to be adequate for detecting polyps >5 mm [28-30]. The endoscopists were blinded to the study groups. Secondary end points were adenoma detection rate (ADR), polyp detection rate (PDR), cecal intubation time, and withdrawal time. ADR and PDR were calculated by dividing the number of patients with at least one adenoma and one polyp, respectively, by the total number of colonoscopy patients (based on the histological diagnosis according to the revised Vienna classification) [19,20]. Withdrawal time included the time from starting withdrawal from the cecum to the final inspection of the rectum, including the time spent on washing, suctioning, and polypectomies.

Items from the PSQ-18 were transformed to bowel preparation education purposes to investigate patient satisfaction [26]. Scores for the following subscales were calculated by averaging the scores of the relevant questions: general satisfaction (items 3 and 6), technical quality (items 8 and 9), communication (items 1 and 2), time spent on education (item 7), and convenience (items 4 and 5). Responses to all items were given on a five-point Likert scale, ranging from strongly agree to strongly disagree. Patients in the smartphone app group were also asked to rate the user friendliness and design of the smartphone app on a 10-point scale.

Statistical Analysis and Sample Size

Sample size calculation was performed using PS Power and Sample Size Program version 3.1.2 (W D Dupont and W D Plummer, Jr). To detect a difference of 0.75 in the total BBPS scores between both groups with a significance level (*P* value) of .05 and a power of 80%, 82 completers per group were needed [21,22]. To account for patients dropping out, 90 patients per group were enrolled.

Intention-to-treat analyses were performed. Descriptive statistics are presented as mean (SD) or as the number of patients (%). Differences between study groups were analyzed using two-tailed independent-samples *t* test for numerical variables and chi-square test or Fisher exact test for categorical variables. Posthoc analyses were performed for subgroup analyses. Two-sided *P* values ≤.05 were considered statistically significant. Statistical analyses were performed using SPSS Statistics for Windows, version 25 (IBM).

Results

Study Population

Patients who underwent a colonoscopy at the Maastricht University Medical Center+ between August 2018 and November 2019 were screened for eligibility. In total, 90 patients were included in the smartphone app group and 90 in the control group (Figure 2). A total of 7 patients were excluded from the study. Patient characteristics are provided in Table 1. No significant differences were observed between the smartphone app group and the control group in terms of baseline characteristics. Patients in both the groups had the same level of experience in using medical smartphone apps.

Figure 2. Study flowchart of patient enrollment and inclusion.

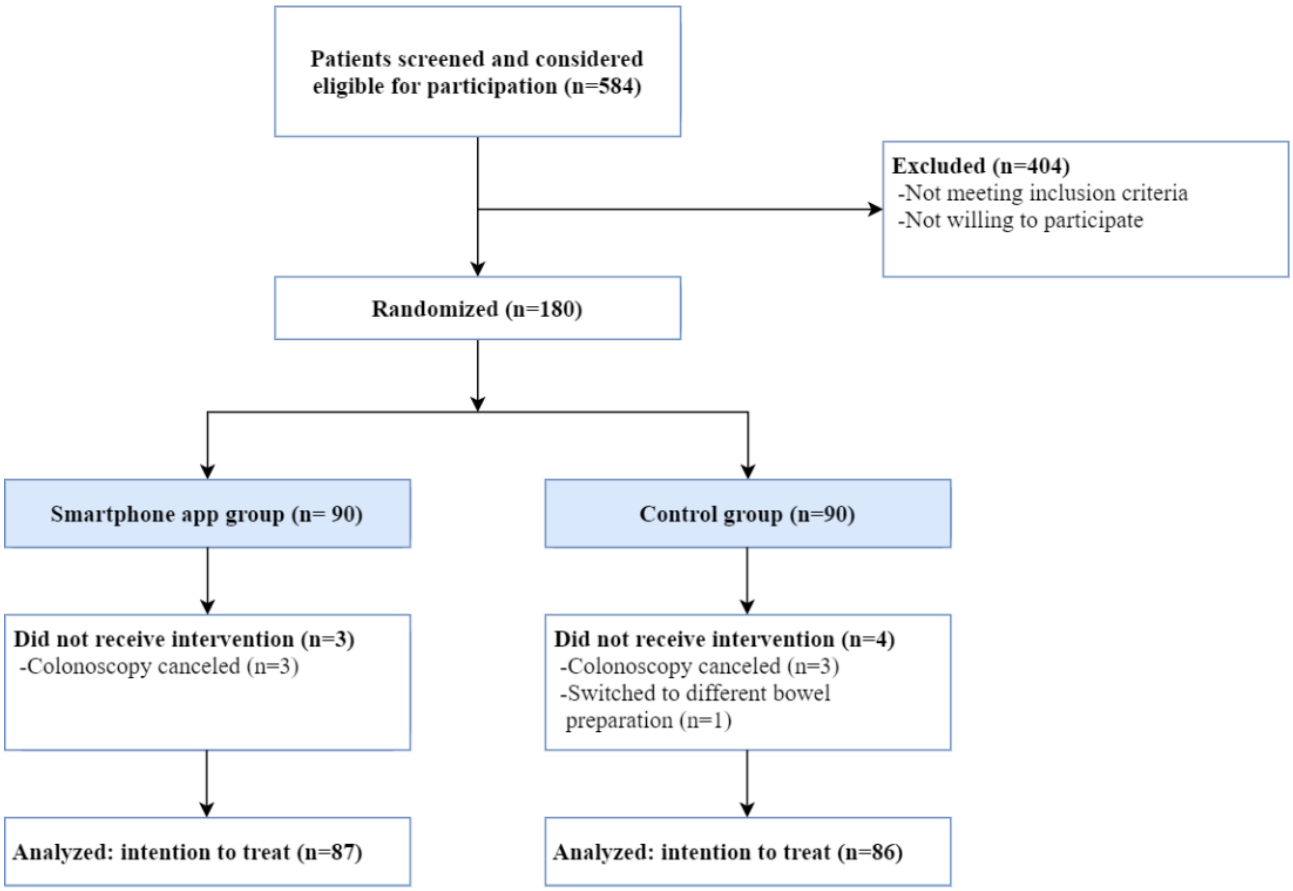


Table 1. Baseline characteristics of patients in the smartphone app group and patients in the control group.

Baseline characteristics	Smartphone app group (N=87)	Control group (N=86)	*P* value
Age (years)			
Value, mean (SD)	56.9 (10.8)	57.1 (12.4)	.92
Age<65, n (%)	67 (77)	62 (72)	.46
Age≥65, n (%)	20 (23)	24 (28)	.46
Gender, female, n (%)	37 (43)	34 (40)	.69
BMI in kg/m^2, mean (SD)	26.1 (4.6)	25.7 (3.6)	.56
Indication for colonoscopy, n (%)			.25
National screening program	29 (33)	21 (24)	
Surveillance	17 (20)	25 (29)	
Symptoms	41 (47)	40 (47)	
Waiting time in days, mean (SD)[a]	26.8 (17.6)	31.6 (24.6)	.14
Previous colonoscopy, n (%)	34 (39)	37 (43)	.60
Gastrointestinal history, n (%)[b]	37 (43)	43 (50)	.32
Diverticulosis	10 (11)	16 (19)	.19
Constipation	14 (16)	18 (21)	.41
Abdominal or pelvic surgery[c]	22 (25)	16 (19)	.28
Comorbidities, n (%)[d]	45 (52)	37 (43)	.25
Level of education, n (%)			.37
High school	15 (20)	9 (13)	
Secondary vocational education	28 (37)	24 (34)	
Higher education (including Bachelor and Master programs at universities of applied sciences)	33 (43)	38 (54)	
Experienced in using smartphone apps, n (%)	76 (99)	59 (86)	.003
More than 10 apps	52 (69)	43 (73)	.65
Previous medical smartphone app use, n (%)	7 (9)	8 (12)	.62

[a]Waiting time was defined as the time between screening visit and colonoscopy.

[b]Inflammatory bowel disease and stenosis did not occur in any patients' medical history.

[c]Abdominal or pelvic surgery included colectomy, abdominal uterus extirpation, prostatectomy, appendectomy, nephrectomy, cholecystectomy, and cesarean delivery.

[d]Comorbidities included hypertension, cardiovascular disease, chronic pulmonary disease, renal disease, liver disease, psychiatric disease, and diabetes mellitus.

Bowel Preparation Quality

Colonoscopies were performed by 25 different endoscopists (gastroenterologists and fellows) who rated the BBPS. All endoscopists were experienced in scoring the BBPS. The mean total BBPS score in the smartphone app group was significantly higher than that in the control group (mean 8.3, SD 0.9 vs mean 7.9, SD 1.2; *P*=.03). Mean right colon segment scores were also significantly higher in the smartphone app group (mean 2.7, SD 0.5 vs mean 2.5, SD 0.6; *P*=.04). No significant differences were observed in the mean transverse colon and left colon segment scores (Table 2). One patient in the smartphone app group and 4 patients in the control group had inadequate bowel preparation scores (*P*=.18). Multivariable logistic regression analyses, to reveal independent predictors for inadequate bowel preparation, could not be performed because of this low number.

Table 2. Bowel preparation scores for the smartphone app group and the control group.[a]

Bowel preparation quality	Smartphone app group (n=81)	Control group (n=81)	*P* value
BBPS,[b] mean (SD)			
Total	8.3 (0.9)	7.9 (1.2)	*.03* [c]
BBPS right colon	2.7 (0.5)	2.5 (0.6)	*.04*
BBPS transverse colon	2.8 (0.4)	2.8 (0.4)	.34
BBPS left colon	2.8 (0.4)	2.6 (0.5)	.07
Adequate bowel preparation, n (%)[d]	80 (99)	77 (95)	.18[e]
Total BBPS score ≥6	81 (100)	79 (98)	.25[e]
All segment scores ≥2	80 (99)	77 (95)	.18[e]

[a]Analyses for the Boston Bowel Preparation Scale included only complete colonoscopies (successful cecal intubation). Missing data were equally distributed between the smartphone app group (n=5) and the control group (n=5). Analyses including incomplete colonoscopies showed similar results.

[b]BBPS: Boston Bowel Preparation Scale.

[c]Italicization represents statistically significant result ($P<.05$).

[d]Adequate bowel preparation was defined as a total Boston Bowel Preparation Scale score of ≥6 and segment scores of ≥2.

[e]Fisher exact test.

Subgroup analyses were performed for morning and afternoon colonoscopies, age below and above 65 years, and colonoscopy waiting time exceeding 1 month or not (because of an increased risk of forgetting preparation instructions over time; Table 3). These analyses showed that patients aged <65 years in the smartphone app group had a significantly higher mean total (mean 8.4, SD 0.9 vs mean 7.9, SD 1.1; *P*=.01) and right BBPS score (mean 2.8, SD 0.4 vs mean 2.5, SD 0.6; *P*=.01) than those in the control group. Patients in the smartphone app group having an afternoon colonoscopy also had a significantly higher mean total and right BBPS score than those in the control group. Furthermore, patients with a colonoscopy waiting time >1 month in the smartphone app group had a significantly higher mean total BBPS score and a significantly cleaner left colon than those in the control group. No significant differences were observed for morning colonoscopies, age ≥65 years, and colonoscopies performed within 1 month.

Table 3. Subgroup analysis for the smartphone app group and the control group.[a]

Subgroup analyses	Smartphone app group (n=81)	Control group (n=81)	*P* value
Afternoon colonoscopy			
Patient, n (%)	37 (46)	35 (43)	.75[b]
Total BBPS,[c] mean (SD)	8.3 (1.0)	7.7 (1.3)	.03
BBBS right colon, mean (SD)	2.7 (0.5)	2.4 (0.6)	.04
BBPS transverse colon, mean (SD)	2.8 (0.4)	2.7 (0.5)	.50
BBPS left colon, mean (SD)	2.8 (0.4)	2.5 (0.6)	.02
Morning colonoscopy			
Patient, n (%)	44 (54)	46 (57)	.75[b]
Total BBPS, mean (SD)	8.3 (0.9)	8.1 (1.0)	.37
BBBS right colon, mean (SD)	2.7 (0.5)	2.6 (0.5)	.38
BBPS transverse colon, mean (SD)	2.8 (0.4)	2.8 (0.4)	.49
BBPS left colon, mean (SD)	2.8 (0.4)	2.7 (0.5)	.73
Age <65 years			
Patient, n (%)	63 (78)	57 (70)	.28[b]
Total BBPS, mean (SD)	8.4 (0.9)	7.9 (1.1)	.01
BBBS right colon, mean (SD)	2.8 (0.4)	2.5 (0.6)	.01
BBPS transverse colon, mean (SD)	2.8 (0.4)	2.7 (0.4)	.17
BBPS left colon, mean (SD)	2.8 (0.4)	2.7 (0.5)	.14
Age ≥65 years			
Patient, n (%)	18 (22)	24 (30)	.28[b]
Total BBPS, mean (SD)	7.9 (1.0)	7.9 (1.3)	.94
BBBS right colon, mean (SD)	2.4 (0.6)	2.5 (0.6)	.61
BBPS transverse colon, mean (SD)	2.7 (0.5)	2.8 (0.4)	.61
BBPS left colon, mean (SD)	2.7 (0.5)	2.6 (0.5)	.37
Colonoscopy waiting time >1 month			
Patient, n (%)	27 (33)	36 (44)	.15[b]
Total BBPS, mean (SD)	8.3 (0.8)	7.7 (1.1)	.02
BBBS right colon, mean (SD)	2.6 (0.6)	2.4 (0.6)	.31
BBPS transverse colon, mean (SD)	2.9 (0.4)	2.7 (0.5)	.21
BBPS left colon, mean (SD)	2.9 (0.4)	2.5 (0.5)	.004
Colonoscopy waiting time ≤1 month			
Patient, n (%)	54 (67)	45 (56)	.15[b]
Total BBPS, mean (SD)	8.3 (1.0)	8.1 (1.2)	.38
BBBS right colon, mean (SD)	2.7 (0.5)	2.6 (0.6)	.12
BBPS transverse colon, mean (SD)	2.8 (0.4)	2.8 (0.4)	.83
BBPS left colon, mean (SD)	2.7 (0.4)	2.7 (0.5)	.94

[a]Analyses for the Boston Bowel Preparation Scale included only complete colonoscopies (successful cecal intubation). Missing data were equally distributed between the smartphone app group (n=6) and the control group (n=5). Analyses including incomplete colonoscopies showed similar results.

[b]Chi-square test comparing presence in specific subgroups (afternoon vs morning, age <65 years vs age ≥65 years, and colonoscopy waiting time ≤1 month vs colonoscopy waiting time >1 month) between the smartphone app group and the control group.

[c]BBPS: Boston Bowel Preparation Scale.

Colonoscopy Quality Parameters

The cecal intubation rate was 93% and 94% in the smartphone app group and the control group, respectively (*P*=.77; Table 4). Eleven colonoscopies were incomplete because of severe pain sensations (n=6), stenosis (n=3), and technical difficulties (n=2). No colonoscopies were aborted because of inadequate bowel preparation. The mean withdrawal time did not differ significantly between the smartphone app group and the control group (Table 4). Both ADR and PDR were higher in patients in the smartphone app group than in patients in the control group, but the difference was not statistically significant.

Table 4. Colonoscopy quality parameters for the smartphone app group and the control group.

Colonoscopy quality parameters	Smartphone app group (n=87)	Control group (n=86)	*P* value
Cecal intubation rate, n (%)	81 (93)	81 (94)	.77
Withdrawal time in minutes, mean (SD)[a]	15.8 (8.6)	14.0 (9.1)	.20
Adenoma detection rate, n (%)[b]	35 (43)	27 (33)	.20
Polyp detection rate, n (%)[b]	44 (54)	36 (44)	.20

[a]Analyses for withdrawal time included only complete colonoscopies (successful cecal intubation). Withdrawal time could not be calculated for n=3 in the smartphone app group and not for n=1 in the control group.

[b]Analyses for adenoma and polyp detection rate included only complete colonoscopies (successful cecal intubation). Missing data were equally distributed between the smartphone app group (n=6) and the control group (n=5). Analyses including incomplete colonoscopies showed similar results.

Patient Satisfaction

The response rates of the PSQ-18 were 85% (74/87) in the smartphone app group and 83% (71/86) in the control group (*P*=.66). On a five-point Likert scale, the general satisfaction was 4.4 (SD 0.7) in the smartphone app group and 4.3 (SD 0.8) in the control group (*P*=.32). No significant differences in patient satisfaction were observed in terms of technical quality, communication, time spent on education, and convenience (Table 5). The majority of smartphone app users were willing to use the app again for eventual future colonoscopies (mean 4.5, SD 0.6) and rated the added value of the smartphone app 4.4 (SD 0.7). On a 10-point scale, user friendliness and design of the smartphone app were rated 8.7 (SD 1.1) and 8.7 (SD 1.2), respectively.

Table 5. Patient satisfaction according to the Patient Satisfaction Questionnaire-18 and patient satisfaction with smartphone app use.[a]

Patient satisfaction	Smartphone app group (n=74)	Control group (n=71)	*P* value
PSQ-18[b] (5-point scale), mean (SD)			
General satisfaction	4.4 (0.7)	4.3 (0.8)	.32
Technical quality	4.5 (0.7)	4.5 (0.6)	.70
Communication	4.6 (0.5)	4.7 (0.6)	.52
Time spent on education	4.6 (0.7)	4.7 (0.6)	.45
Convenience	4.4 (0.7)	4.5 (0.6)	.45
Patient satisfaction on smartphone app use (5-point scale),[c] mean (SD)			
Added value of the smartphone app	4.4 (0.7)	N/A[d]	N/A
Willingness to use the app for future colonoscopies	4.5 (0.6)	N/A	N/A
Ease of downloading and using	4.6 (0.7)	N/A	N/A
Clear overview of times to use laxative	4.6 (0.7)	N/A	N/A
Patient satisfaction on smartphone app use (10-point scale),[c] mean (SD)			
Ease of use in general	8.7 (1.1)	N/A	N/A
Design	8.7 (1.2)	N/A	N/A

[a]Analyses for patient satisfaction included only complete questionnaires. Analyses including incomplete questionnaires showed similar results.

[b]PSQ-18: Patient Satisfaction Questionnaire-18.

[c]Analyses for patient satisfaction on smartphone app use was only applicable for smartphone app users and based on n=78 complete questionnaires.

[d]N/A: not applicable.

Discussion

Principal Findings

Adequate bowel preparation is an important quality indicator for colonoscopy. The key finding of this study is the significantly higher mean total BBPS score in patients using a personalized smartphone app for bowel preparation instructions compared with patients using regular verbal and written information. Patient satisfaction did not improve further for smartphone app users compared with patients receiving regular written instructions.

Comparison With Previous Work

The finding of a significantly higher mean total BBPS score in the smartphone app group compared with the control group is in line with previous studies [2,31,32]. The mean total BBPS score in the control groups of these studies ranged from 5.8 to 7.2. Although the mean total BBPS score (mean 7.9, SD 1.2) in our control group was high, the smartphone app still had added value (mean total BBPS score 8.3, SD 0.9). In particular, the mean BBPS score of the right colon was significantly higher in the smartphone app group than in the control group. This finding is clinically relevant because the right colon is considered more difficult to clean [33] and the PDR in the right colon improves with improvement in BBPS score of the right colon [34].

The European Society of Gastrointestinal Endoscopy recommends the use of enhanced instructions for bowel preparation. Methods such as telephone calls, visual aids, educational videos, and SMS reminders help to improve bowel preparation quality compared with regular instructions [1,2,4,13,33,35-37]. Possible advantages of smartphone apps are that they are more easily understandable, accessible, and interactive. Another benefit is that automatic alerts, reminders, and notifications remind patients to start and adhere to the steps of the bowel preparation schedule more precisely [38,39] without consuming valuable time and resources, as is the case with telephone calls [13,35], making smartphone apps easier to implement in daily clinical practice. Furthermore, the smartphone app provided a personalized bowel preparation schedule for each patient. The different steps of the bowel preparation procedure were adapted to the exact date and time of colonoscopy. In contrast, written instructions were general for morning and afternoon colonoscopies and indicated no exact date.

Previous studies included relatively young patients with a mean age of 42-55 years [2,21,35,36]. In this study, no maximum age for participation was stated, so older age groups, who might be less familiar with smartphone apps, were also included. Jeon et al [37] used a smartphone mobile messenger to educate patients and found that this approach was useful with respect to the quality of bowel preparation for the younger age group (<40 years) but not for patients aged >40 years. In our study, subgroup analysis showed significantly higher total mean BBPS scores and right colon segment scores for patients aged <65 years using the smartphone app compared with the control group. In addition to the study by Jeon et al [37], the significantly higher mean BBPS scores indicate that the use of a smartphone app is a feasible method not only for patients aged <40 years but also for patients aged <65 years. For patients aged ≥65 years, no significant differences in mean BBPS scores were found, although their number was low. Further research focusing on older patients (≥65 years) is needed to investigate the usefulness of a smartphone app among these patients.

In this study, the BBPS was used to measure bowel cleansing. A systematic review by Parmar et al [27] revealed that the BBPS is the most thoroughly validated scale and should therefore be used in clinical practice. It should be noted that the BBPS is scored after appropriate washing and suctioning steps have been performed. Therefore, differences in initial bowel preparation could have been masked by variations in the extent of the endoscopists' washing and suctioning actions. However, because blinded endoscopists performed colonoscopies in both groups, potential differences in the extent of washing and suctioning were eliminated.

The minimum standard rate for adequate bowel preparation of ≥90%, a set criterion by the European Society of Gastrointestinal Endoscopy guidelines [40], was reached in both the smartphone app group and the control group. In 5 patients (5/173, 2.9%), the colon was inadequately prepared. In the literature, the reported numbers are higher, up to 35% [7,10,13,35]. In this study, predictors for inadequate bowel preparation could not be identified because of the low number of patients. In two meta-analyses, three groups of predictors for inadequate bowel preparation were identified: patients' characteristics (increasing age, male gender, and higher BMI), clinical conditions (constipation, diabetes mellitus, hypertension, cirrhosis, stroke, and dementia), and medication use (narcotics and tricyclic antidepressants) [41,42]. Other studies also reported low level of education, low socioeconomic status, low health literacy, and low patient motivation in health promotion as influencing factors [13,35].

ADR and cecal intubation rate are indicators of colonoscopy quality [30]. Guo et al [43] found a significantly higher ADR in the smartphone app group than in the control group (21.4% vs 12.8%, respectively; P=.03). Although higher ADR and PDR were observed in the smartphone app group in this study, the observed differences were not statistically significant. It should be noted that this study was not powered to detect significant differences in ADR and PDR. A recent meta-analysis found that patients who had received enhanced instructions (social media apps, SMS, and telephone calls) had higher cecal intubation rates (odds ratio 2.77, 95% CI 1.73-4.42; P<.001) than patients receiving regular verbal and written instructions [4]. In this study, none of the cases in which the cecum was not reached were because of inadequate bowel preparation, although it has been reported as a major factor in the literature [44].

Bowel preparation procedures may cause discomfort. The main discomfort patients report relates to uncertainties with respect to dietary recommendations and adverse gastrointestinal symptoms owing to use of laxatives [33]. Patient education using a smartphone app may help resolve these uncertainties [45]. Indeed, the willingness to repeat the preparation procedure was higher for patients receiving enhanced bowel preparation instructions than for those receiving regular instructions (odds

ratio 1.91, 95% CI 1.20-3.04; *P*=.01) [4]. High patient satisfaction can therefore help to increase patient participation in surveillance colonoscopies. In our control group, patient satisfaction was already high and increased further when using the smartphone app.

Strengths and Limitations

This study had several strengths. Selection bias was avoided in three ways. First, inclusion concerned screening, surveillance, and symptomatic patients of both morning and afternoon colonoscopies. Second, patients were not excluded if they had a history of abdominal surgery, diverticulosis, stenosis, or constipation, compared with most other studies [2,22,31,35,46]. Third, the app was available for smartphones with both Android and iOS operating systems, in contrast to the study by Lorenzo-Zuniga et al [36]. Furthermore, no maximum age for participation was stated. All the abovementioned decisions in the methodology add to the generalizability of our findings.

This study also had certain limitations. First, compliance with the bowel preparation schedule was not controlled in either group, although it is known that approximately 30% of patients with poor bowel preparation fail to follow instructions before the colonoscopy [23]. In addition, we did not monitor other variables related to BBPS, such as searching for additional information on the internet or other social media or help provided by other sources or people. Second, the patients were not blinded to the intervention. Third, a large number of endoscopists assessed the BBPS, potentially leading to a larger variability in scoring and possibly causing bias. All endoscopists were trained and experienced in using the BBPS to achieve uniform scoring, thereby reflecting daily endoscopic practice in a teaching hospital. Fourth, selection bias may have occurred, as only 30.8% (180/584) of the screened patients visiting our prescreen facility were eligible for inclusion. Most likely, only patients with an affinity for smartphone use were willing to participate, lowering the generalizability of this study. With the expectation of an increase in smartphone use in the future, generalizability will subsequently increase, and smartphone apps for bowel preparation can be a valuable tool in improving bowel preparation quality. Fifth, the study was performed at a single center, limiting its generalizability.

Conclusions

In conclusion, this study showed that using our personalized smartphone app significantly improved bowel preparation quality, particularly in the right colon, and could improve polyp detection in the right colon. Patient satisfaction was equal in the personalized smartphone app group and the control group. Smartphone apps are an easy-to-use tool to improve patients' bowel preparation education and quality, making implementation in clinical practice feasible.

Acknowledgments

Ferring B V financially supported and facilitated the development of the smartphone app (Prepit, Prepare for colonoscopy, Ferring B V). Ferring B V did not contribute to the study protocol, data analysis, or manuscript writing.

Authors' Contributions

QEWVDZ, AR, BVDV, AAMM, and RJJDR conceptualized the study and its design, drafted the paper, analyzed and interpreted the data, critically revised the paper for important intellectual content, and approved the final paper. BW analyzed and interpreted the data, critically revised the paper for important intellectual content, and approved the final paper.

Conflicts of Interest

AAMM was supported by a health care efficiency grant from ZonMw (Organization for Health Research and Development, the Netherlands), an unrestricted research grant from Will Pharma S A, a restricted educational grant from Ferring B V, and research funding from Allegan and Grünenthal; provided scientific advice to Bayer, Kyowa Kirin, and Takeda; and received a research grant from PENTAX Europe GmbH and the Dutch Cancer Society. RJJDR was supported by a restricted educational grant from Ferring B V. QEWVDZ, AR, BVDV, and BW declare no conflicts of interest for this paper. Ferring B V financially supported and facilitated the development of the smartphone app. Ferring B V had no role in the design, practice, or analysis of this study.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1221 KB - mhealth_v9i8e26703_app1.pdf]

Multimedia Appendix 2
eHealth checklist patient satisfaction questionnaire-18.
[DOCX File , 16 KB - mhealth_v9i8e26703_app2.docx]

References

1. Hassan C, East J, Radaelli F, Spada C, Benamouzig R, Bisschops R, et al. Bowel preparation for colonoscopy: European Society of Gastrointestinal Endoscopy (ESGE) Guideline - Update 2019. Endoscopy 2019 Aug;51(8):775-794 [FREE Full text] [doi: 10.1055/a-0959-0505] [Medline: 31295746]

2. Cho J, Lee S, Shin JA, Kim JH, Lee HS. The impact of patient education with a smartphone application on the quality of bowel preparation for screening colonoscopy. Clin Endosc 2017 Sep;50(5):479-485 [FREE Full text] [doi: 10.5946/ce.2017.025] [Medline: 28669148]
3. Clark BT, Rustagi T, Laine L. What level of bowel prep quality requires early repeat colonoscopy: systematic review and meta-analysis of the impact of preparation quality on adenoma detection rate. Am J Gastroenterol 2014 Nov;109(11):1714-1724 [FREE Full text] [doi: 10.1038/ajg.2014.232] [Medline: 25135006]
4. Guo X, Yang Z, Zhao L, Leung F, Luo H, Kang X, et al. Enhanced instructions improve the quality of bowel preparation for colonoscopy: a meta-analysis of randomized controlled trials. Gastrointest Endosc 2017 Jan;85(1):90-97. [doi: 10.1016/j.gie.2016.05.012] [Medline: 27189659]
5. Chokshi RV, Hovis CE, Hollander T, Early DS, Wang JS. Prevalence of missed adenomas in patients with inadequate bowel preparation on screening colonoscopy. Gastrointest Endosc 2012 Jun;75(6):1197-1203. [doi: 10.1016/j.gie.2012.01.005] [Medline: 22381531]
6. le Clercq CM, Bouwens MW, Rondagh EJ, Bakker CM, Keulen ET, de Ridder RJ, et al. Postcolonoscopy colorectal cancers are preventable: a population-based study. Gut 2014 Jun;63(6):957-963. [doi: 10.1136/gutjnl-2013-304880] [Medline: 23744612]
7. Harewood GC, Sharma VK, de Garmo P. Impact of colonoscopy preparation quality on detection of suspected colonic neoplasia. Gastrointest Endosc 2003 Jul;58(1):76-79. [doi: 10.1067/mge.2003.294] [Medline: 12838225]
8. Armstrong D, Barkun AN, Chen Y, Daniels S, Hollingworth R, Hunt RH, et al. Access to specialist gastroenterology care in Canada: the Practice Audit in Gastroenterology (PAGE) Wait Times Program. Can J Gastroenterol 2008 Feb;22(2):155-160 [FREE Full text] [doi: 10.1155/2008/292948] [Medline: 18299734]
9. Luck A, Pearson S, Maddern G, Hewett P. Effects of video information on precolonoscopy anxiety and knowledge: a randomised trial. Lancet 1999 Dec 11;354(9195):2032-2035. [doi: 10.1016/s0140-6736(98)10495-6] [Medline: 10636368]
10. Chang C, Shih S, Wang H, Chu C, Wang T, Hung C, et al. Meta-analysis: The effect of patient education on bowel preparation for colonoscopy. Endosc Int Open 2015 Dec 24;3(6):646-652 [FREE Full text] [doi: 10.1055/s-0034-1392365] [Medline: 26716129]
11. Park J, Sohn C, Hwang S, Choi H, Park J, Kim H, et al. Quality and effect of single dose versus split dose of polyethylene glycol bowel preparation for early-morning colonoscopy. Endoscopy 2007 Jul 5;39(7):616-619. [doi: 10.1055/s-2007-966434] [Medline: 17611916]
12. Gimeno-García AZ, de la Barreda Heuser R, Reygosa C, Hernandez A, Mascareño I, Nicolás-Pérez D, et al. Impact of a 1-day versus 3-day low-residue diet on bowel cleansing quality before colonoscopy: a randomized controlled trial. Endoscopy 2019 Jul;51(7):628-636. [doi: 10.1055/a-0864-1942] [Medline: 30943553]
13. Liu Z, Zhang MM, Li YY, Li LX, Li YQ. Enhanced education for bowel preparation before colonoscopy: a state-of-the-art review. J Dig Dis 2017 Feb;18(2):84-91. [doi: 10.1111/1751-2980.12446] [Medline: 28067463]
14. ASGE Standards of Practice Committee, Saltzman JR, Cash BD, Pasha SF, Early DS, Muthusamy VR, et al. Bowel preparation before colonoscopy. Gastrointest Endosc 2015 Apr;81(4):781-794. [doi: 10.1016/j.gie.2014.09.048] [Medline: 25595062]
15. Hassan C, Bretthauer M, Kaminski MF, Polkowski M, Rembacken B, Saunders B, European Society of Gastrointestinal Endoscopy. Bowel preparation for colonoscopy: European Society of Gastrointestinal Endoscopy (ESGE) guideline. Endoscopy 2013;45(2):142-150 [FREE Full text] [doi: 10.1055/s-0032-1326186] [Medline: 23335011]
16. Global mobile consumer survey 2017 : The Netherlands. Deloitte. URL: https://www2.deloitte.com/content/dam/Deloitte/nl/Documents/technology-media-telecommunications/2017%20GMCS%20Dutch%20Edition.pdf [accessed 2021-07-08]
17. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. J Am Med Assoc 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
18. The General Data Protection Regulation (GDPR). European Parliament and the Council of the European Union. 2016. URL: https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en [accessed 2021-07-08]
19. Schlemper RJ, Riddell RH, Kato Y, Borchard F, Cooper HS, Dawsey SM, et al. The Vienna classification of gastrointestinal epithelial neoplasia. Gut 2000 Aug;47(2):251-255 [FREE Full text] [doi: 10.1136/gut.47.2.251] [Medline: 10896917]
20. Rex DK, Ponugoti PL. Calculating the adenoma detection rate in screening colonoscopies only: is it necessary? Can it be gamed? Endoscopy 2017 Nov;49(11):1069-1074. [doi: 10.1055/s-0043-113445] [Medline: 28753699]
21. Park J, Kim MS, Kim H, Kim SI, Shin CH, Lee HJ, et al. A randomized controlled trial of an educational video to improve quality of bowel preparation for colonoscopy. BMC Gastroenterol 2016 Jun 17;16(1):64 [FREE Full text] [doi: 10.1186/s12876-016-0476-6] [Medline: 27317249]
22. Liu X, Luo H, Zhang L, Leung FW, Liu Z, Wang X, et al. Telephone-based re-education on the day before colonoscopy improves the quality of bowel preparation and the polyp detection rate: a prospective, colonoscopist-blinded, randomised, controlled study. Gut 2014 Jan;63(1):125-130. [doi: 10.1136/gutjnl-2012-304292] [Medline: 23503044]
23. Nguyen DL, Wieland M. Risk factors predictive of poor quality preparation during average risk colonoscopy screening: the importance of health literacy. J Gastrointestin Liver Dis 2010 Dec;19(4):369-372 [FREE Full text] [Medline: 21188326]

24. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
25. prepit. App Store Preview. 2021. URL: https://apps.apple.com/nl/app/prepit/id1204337501 [accessed 2021-07-21]
26. Marshall GN, Hays RD. The Patient Satisfaction Questionnaire Short-Form (PSQ-18). RAND. 1994. URL: https://www.rand.org/content/dam/rand/pubs/papers/2006/P7865.pdf [accessed 2021-07-08]
27. Parmar R, Martel M, Rostom A, Barkun AN. Validated scales for colon cleansing: a systematic review. Am J Gastroenterol 2016 Feb;111(2):197-205. [doi: 10.1038/ajg.2015.417] [Medline: 26782820]
28. Calderwood AH, Schroy PC, Lieberman DA, Logan JR, Zurfluh M, Jacobson BC. Boston Bowel Preparation Scale scores provide a standardized definition of adequate for describing bowel cleanliness. Gastrointest Endosc 2014 Aug;80(2):269-276 [FREE Full text] [doi: 10.1016/j.gie.2014.01.031] [Medline: 24629422]
29. Clark BT, Protiva P, Nagar A, Imaeda A, Ciarleglio MM, Deng Y, et al. Quantification of adequate bowel preparation for screening or surveillance colonoscopy in men. Gastroenterology 2016 Feb;150(2):396-405 [FREE Full text] [doi: 10.1053/j.gastro.2015.09.041] [Medline: 26439436]
30. Kastenberg D, Bertiger G, Brogadir S. Bowel preparation quality scales for colonoscopy. World J Gastroenterol 2018 Jul 14;24(26):2833-2843 [FREE Full text] [doi: 10.3748/wjg.v24.i26.2833] [Medline: 30018478]
31. Walter B, Frank R, Ludwig L, Dikopoulos N, Mayr M, Neu B, et al. Smartphone application to reinforce education increases high-quality preparation for colorectal cancer screening colonoscopies in a randomized trial. Clin Gastroenterol Hepatol 2020 Mar 30;19(2):331-338. [doi: 10.1016/j.cgh.2020.03.051] [Medline: 32240835]
32. Wang S, Wang Q, Yao J, Zhao S, Wang L, Li Z, et al. Effect of WeChat and short message service on bowel preparation: an endoscopist-blinded, randomized controlled trial. Eur J Gastroenterol Hepatol 2019 Feb;31(2):170-177. [doi: 10.1097/MEG.0000000000001303] [Medline: 30418256]
33. Walter B, Klare P, Strehle K, Aschenbeck J, Ludwig L, Dikopoulos N, et al. Improving the quality and acceptance of colonoscopy preparation by reinforced patient education with short message service: results from a randomized, multicenter study (PERICLES-II). Gastrointest Endosc 2019 Mar;89(3):506-513. [doi: 10.1016/j.gie.2018.08.014] [Medline: 30138612]
34. Clark BT, Laine L. High-quality bowel preparation is required for detection of sessile serrated polyps. Clin Gastroenterol Hepatol 2016 Aug;14(8):1155-1162 [FREE Full text] [doi: 10.1016/j.cgh.2016.03.044] [Medline: 27060426]
35. Sharara AI, Chalhoub JM, Beydoun M, Shayto RH, Chehab H, Harb AH, et al. A customized mobile application in colonoscopy preparation: a randomized controlled trial. Clin Transl Gastroenterol 2017 Jan 05;8(1):e211 [FREE Full text] [doi: 10.1038/ctg.2016.65] [Medline: 28055031]
36. Lorenzo-Zúñiga V, Moreno DV, Marín I, Barberá M, Boix J. Improving the quality of colonoscopy bowel preparation using a smart phone application: a randomized trial. Dig Endosc 2015 Jul;27(5):590-595. [doi: 10.1111/den.12467] [Medline: 25708251]
37. Jeon SC, Kim JH, Kim SJ, Kwon HJ, Choi YJ, Jung K, et al. Effect of sending educational video clips via smartphone mobile messenger on bowel preparation before colonoscopy. Clin Endosc 2019 Jan;52(1):53-58 [FREE Full text] [doi: 10.5946/ce.2018.072] [Medline: 30153724]
38. Desai M, Nutalapati V, Bansal A, Buckles D, Bonino J, Olyaee M, et al. Use of smartphone applications to improve quality of bowel preparation for colonoscopy: a systematic review and meta-analysis. Endosc Int Open 2019 Feb 18;7(2):216-224 [FREE Full text] [doi: 10.1055/a-0796-6423] [Medline: 30705956]
39. Lee YJ, Kim ES, Choi JH, Lee KI, Park KS, Cho KB, et al. Impact of reinforced education by telephone and short message service on the quality of bowel preparation: a randomized controlled study. Endoscopy 2015 Nov;47(11):1018-1027. [doi: 10.1055/s-0034-1392406] [Medline: 26182387]
40. Kaminski MF, Thomas-Gibson S, Bugajski M, Bretthauer M, Rees CJ, Dekker E, et al. Performance measures for lower gastrointestinal endoscopy: a European Society of Gastrointestinal Endoscopy (ESGE) Quality Improvement Initiative. Endoscopy 2017 Apr;49(4):378-397 [FREE Full text] [doi: 10.1055/s-0043-103411] [Medline: 28268235]
41. Gandhi K, Tofani C, Sokach C, Patel D, Kastenberg D, Daskalakis C. Patient characteristics associated with quality of colonoscopy preparation: a systematic review and meta-analysis. Clin Gastroenterol Hepatol 2018 Mar;16(3):357-369. [doi: 10.1016/j.cgh.2017.08.016] [Medline: 28826680]
42. Mahmood S, Farooqui SM, Madhoun MF. Predictors of inadequate bowel preparation for colonoscopy: a systematic review and meta-analysis. Eur J Gastroenterol Hepatol 2018 Aug;30(8):819-826. [doi: 10.1097/MEG.0000000000001175] [Medline: 29847488]
43. Guo B, Zuo X, Li Z, Liu J, Xu N, Li X, et al. Improving the quality of bowel preparation through an app for inpatients undergoing colonoscopy: a randomized controlled trial. J Adv Nurs 2020 Apr;76(4):1037-1045. [doi: 10.1111/jan.14295] [Medline: 31840286]
44. Aslinia F, Uradomo L, Steele A, Greenwald BD, Raufman J. Quality assessment of colonoscopic cecal intubation: an analysis of 6 years of continuous practice at a university hospital. Am J Gastroenterol 2006 Apr;101(4):721-731. [doi: 10.1111/j.1572-0241.2006.00494.x] [Medline: 16494586]
45. Liu Y, Geng Z, Wu F, Yuan C. Developing "Information Assistant": a smartphone application to meet the personalized information needs of women with breast cancer. Stud Health Technol Inform 2017;245:156-160. [Medline: 29295072]

46. Back SY, Kim HG, Ahn EM, Park S, Jeon SR, Im HH, et al. Impact of patient audiovisual re-education via a smartphone on the quality of bowel preparation before colonoscopy: a single-blinded randomized study. Gastrointest Endosc 2018 Mar;87(3):789-799. [doi: 10.1016/j.gie.2017.09.007] [Medline: 28939500]

Abbreviations

ADR: adenoma detection rate
BBPS: Boston Bowel Preparation Scale
PDR: polyp detection rate
PSQ-18: Patient Satisfaction Questionnaire-18
SPMC: sodium picosulfate with magnesium citrate

Edited by L Buis; submitted 22.12.20; peer-reviewed by Q Wang, C Ochoa-Zezzatti; comments to author 03.03.21; revised version received 07.04.21; accepted 16.06.21; published 19.08.21.

Please cite as:
van der Zander QEW, Reumkens A, van de Valk B, Winkens B, Masclee AAM, de Ridder RJJ
Effects of a Personalized Smartphone App on Bowel Preparation Quality: Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(8):e26703
URL: https://mhealth.jmir.org/2021/8/e26703
doi:10.2196/26703
PMID:34420924

Original Paper

Supporting Women Undergoing IVF Treatment With Timely Patient Information Through an App: Randomized Controlled Trial

Thomas Timmers[1,2], MSc; Manouk Keijsers[3], MD; Jan A M Kremer[1], MD, PhD; Loes Janssen[4], PhD; Jesper Smeenk[3], MD, PhD

[1]IQ healthcare, Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, Netherlands

[2]Interactive Studios, Rosmalen, Netherlands

[3]Department of Obstetrics and Gynaecology, Elisabeth-TweeSteden Hospital, Tilburg, Netherlands

[4]Máxima Medical Center, Veldhoven, Netherlands

Corresponding Author:
Thomas Timmers, MSc
IQ healthcare
Radboud Institute for Health Sciences
Radboud University Medical Center
Geert Grooteplein Zuid 10
Nijmegen, 6525 GA
Netherlands
Phone: 31 24 3615305
Fax: 31 24 3615300
Email: thomas@interactivestudios.nl

Abstract

Background: Since the introduction of assisted reproductive technologies in 1978, over 2 million in vitro fertilization (IVF) babies have been born worldwide. Patients play a vital role in the success of this treatment. They are required to take fertility medication (hormone injections) to activate the ovaries to produce a sufficient number of oocytes. Later, they need to take medication to increase the chance of the embryo surviving inside the uterus. Patients are educated during an intake consultation at the start of the treatment to minimize the emotional burden and reduce noncompliance. The consultation lasts about 30 to 45 minutes and covers all essential subjects. Even though ample time and energy is spent on patient education, patients still feel anxious, unknowledgeable, and unsupported. As such, electronic health utilizing a smartphone or tablet app can offer additional support, as it allows health care professionals to provide their patients with the correct information at the right time by using push notifications.

Objective: This randomized controlled trial aimed to evaluate the capacity of an app to support IVF patients throughout the different phases of their treatment and assess its effectiveness. The study's primary outcome was to determine the patients' level of satisfaction with the information provided. The secondary outcomes included their level of knowledge, ability to administer the medication, overall experienced quality of the treatment, health care consumption, and app usage.

Methods: This study was performed at a specialized fertility clinic of the nonacademic teaching hospital Elisabeth-TweeSteden Ziekenhuis in Tilburg, the Netherlands. Patients who were scheduled for IVF or intracytoplasmic sperm injection treatments between April 2018 and August 2019 were invited to participate in a physician-blinded, randomized controlled trial.

Results: In total, 54 patients participated (intervention group: n=29). Patients in the intervention group demonstrated a higher level of satisfaction on a 0 to 10 scale (mean 8.43, SD 1.03 vs mean 7.70, SD 0.66; P=.004). In addition, they were more knowledgeable about the different elements of the treatment on a 7 to 35 scale (mean 27.29, SD 2.94 vs mean 23.05, SD 2.76; P<.001). However, the difference disappeared over time. There were no differences between the two patient groups on the other outcomes. In total, 25 patients in the intervention group used the app 1425 times, an average of 57 times per patient.

Conclusions: Our study demonstrates that, in comparison with standard patient education, using an app to provide patients with timely information increases their level of satisfaction. Furthermore, using the app leads to a higher level of knowledge about the steps and procedures of IVF treatment. Finally, the app's usage statistics demonstrate patients' informational needs and their willingness to use an electronic health application as part of their treatment.

Trial Registration: Netherlands Trial Register (NTR) 6959; https://www.trialregister.nl/trial/6959

(JMIR Mhealth Uhealth 2021;9(8):e28104) doi:10.2196/28104

KEYWORDS

patient education; fertilization in vitro; mobile health; health literacy; gynecology

Introduction

Background

Since the introduction of assisted reproductive technologies in 1978, over 2 million IVF (in vitro fertilization) babies have been born worldwide [1]. The technique offers infertile couples the chance to become pregnant and is currently applied over a million times annually in the United States and Europe [2-4]. An IVF or ICSI (intracytoplasmic sperm injection) treatment has many different stages and can easily take up to several months. First, there is the collection of oocytes (mature egg cells) from the ovaries that need to be fertilized by sperm in a lab. After successful fertilization, the oocytes are transferred to the uterus (embryo transfer). Pregnancy then depends, among different factors, on the embryos attaching to the lining of the uterus.

Patients' behavior and adherence to treatment instructions play a vital role in the success of this treatment. First, they are required to take fertility medications (hormone injections) to activate the ovaries to produce a sufficient number of oocytes. Later, they need to take medication to increase the chances of the embryo surviving inside the uterus. The medication comes with very strict regimes in terms of application and timing, and most women suffer from the side effects of using the medications and experience stress related to the treatment process. Patients and clinicians report being anxious, which often results in nonadherence to the treatment process [5,6].

Patients are educated about the process during an intake consultation at the start of the treatment to minimize the emotional burden and reduce the risk of noncompliance. The consultation lasts about 30 to 45 minutes. It covers all the important subjects, including the physiology of the menstrual cycle, administration of the medication (and its side effects), oocyte retrieval and embryo transfers, risks, and the chances of becoming pregnant. Even though ample time and energy is spent on patient education, patients still feel anxious, unknowledgeable, and unsupported [5,7-13]. These emotions often relate to the feeling of being uninformed. In contrast, patients prefer being routinely provided with understandable, structured, and practical information regarding their IVF or ICSI treatments. [9,14-18]. Using eHealth via a smartphone or tablet app allows health care professionals to provide their patients with the right information at the right time through push notifications. These notifications may refer to newly available information, prepare patients for a consultation, or remind patients to take their medication and provide relevant instructions.

Furthermore, the information is readily available, complete, well-structured, and presented in different modes like text and video. It can utilize feedback systems to test and retest patients' understanding of important information. A 2020 systematic review demonstrated the effectiveness of these interventions on many different outcomes, ranging from knowledge and satisfaction to adherence and quality of life [19].

Objectives

This randomized controlled trial aimed to evaluate the capacity of an app to support IVF and ICSI patients throughout the different phases of their treatment and assess its effectiveness. The study's primary outcome was to determine the patients' level of satisfaction with the information provided. The secondary outcomes included their level of knowledge, ability to administer the medication, the overall experienced quality of the treatment, and health care consumption. In addition, app usage statistics were gathered to assess the need for specific information in the app. We hypothesized that providing patients with timely information via an app would positively affect all outcomes compared to standard patient education practices.

Methods

Study Design

This study was performed at a specialized fertility clinic of the nonacademic teaching hospital Elisabeth-TweeSteden Ziekenhuis (ETZ) in Tilburg, the Netherlands. Patients who were scheduled for IVF or ICSI treatment were invited to participate in a physician-blinded, randomized controlled trial between April 2018 and August 2019. The study assessed the effectiveness of an interactive app in addition to the standard care (website and brochures) in a parallel-group design with an equal allocation ratio. The app was used to support and educate patients through the different stages of their treatment, ranging from the intake and medication instructions to the oocyte retrieval, embryo transfer, and pregnancy test. No changes were made to the study design after the study was initiated. We followed the CONSORT (Consolidated Standards of Reporting Trials) guidelines and the CONSORT eHealth checklist [20,21].

Informed Consent and Ethical Considerations

The hospital staff asked patients to consider participating in the study following their first consultation with a fertility physician indicating they were eligible for IVF or ICSI treatment. Interested patients received all the necessary information about the study, and they were offered at least 2 days to reflect on the information. If they had any questions, they could contact the local research coordinator (MK, gynecology resident since 2018) by phone or email. If they agreed to participate in the study, patients signed the informed consent before initiating their treatment. The study was registered at the Netherlands Trial Registry (reference number 6959). The study was approved by the Institutional Review Board of the Maxima Medical Centre (Eindhoven, the Netherlands; reference number N18.030) and the ETZ hospital's local review board.

Participant Selection

Patients scheduled for IVF or ICSI treatments at the ETZ hospital were eligible for inclusion. Additionally, participants were required to be fluent in Dutch and possess an email address and a smartphone or tablet. For the remainder of this article, we will refer to this patient population (patients scheduled for IVF or ICSI treatment) as "IVF patients."

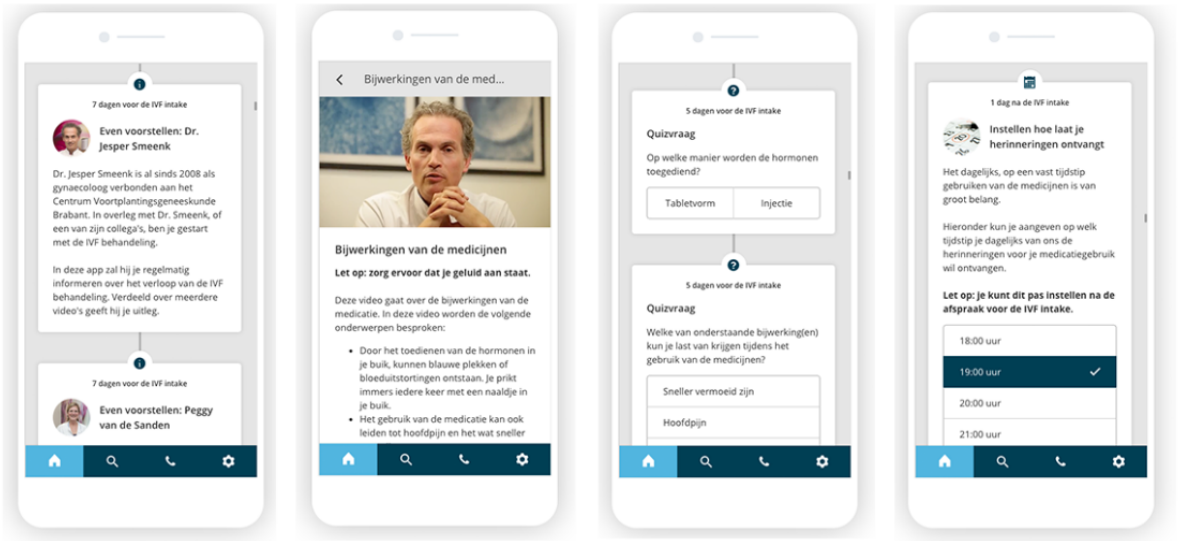
Intervention

The Patient Journey App (Interactive Studios) provided timely information to IVF patients in the intervention group. The app was only available to patients in the intervention group, and they obtained access to the app after completing the baseline questionnaire. They received an email with download instructions for the app and a personal code to enter on the app's timeline to unlock the information in the app.

All patients in the intervention group received the same information via the app. However, the timing of the information and push notifications was based on the date a specific patient started the treatment and the date the patient underwent the oocyte retrieval. These dates were entered into the system by the hospital staff to ensure accuracy. All information, questions, and interactions were provided within the app based on a relative number of days before and after these events. Push notifications were used to alert patients about the newly available information actively. The timing of the push notifications was configured per information item (eg, information about hormone medication side effects was provided 3 days after the intake consultations at 11 am, and information about the preparation for the oocyte puncture appeared 2 days before the oocyte puncture at 8 pm). An overview of the content, notifications, and timing is presented in Multimedia Appendix 1.

The text, photos, and video used in the app were developed specifically for this trial in close collaboration with a gynecologist (JS, subspecialist reproductive medicine since 2011), a clinical embryologist (since 2004), and a specialized fertility care nurse (since 2010). Furthermore, the electronic health records of 10 patients who had previously undergone IVF or ICSI treatment were checked to determine why they had contacted the hospital. Based on this information, an interactive timeline was developed (Figure 1). All information on the timeline was presented in Dutch. No changes were made to the app's content during the trial.

Figure 1. Examples of the interactive app used as an intervention in the study (in Dutch). From left to right: introduction of the different health care providers in the app, video and text information about medication usage and side effects, quiz-like questions to assess patient's knowledge on various topics, and the configuration of patient-specific medication reminders.



Information in the app was tailored to the ETZ hospital and based on existing protocols. Patients that used the app from IVF intake to pregnancy test after a successful embryo transfer received 52 information items and 30 push notifications. The information was disseminated over different phases of the IVF or ICSI process: introduction, welcome to the ETZ fertility center, what is IVF or ICSI, medication usage, IVF or ICSI intake consultation, medication reminders, treatment schedule (hormone injections, side effects, and echography), oocyte retrieval, embryo transfer, and a pregnancy test.

Prior to the study, 4 patients were interviewed to assess the general usefulness and usability of the app. They reported that the app would be very useful and offered no additional suggestions or changes. After the study, all the content developed for the intervention was provided to the fertility clinic, allowing them to offer it to their patients as part of the new standard of care.

Study Outcomes

Patients' satisfaction with the information they received during the treatment was assessed as a primary outcome. Secondary outcomes assessed patients' level of knowledge, satisfaction with the IVF intake consultation, health care consumption, and their ability to understand the information, administer hormone injections, and manage side effects. In addition, we assessed patients' overall satisfaction with the entire treatment process.

Finally, data on app usage was continuously captured to understand better how the app is being used over time, the type of information that patients consult, and the videos they watch (Textbox 1).

Textbox 1. Overview of questionnaires used per outcome.

Outcome and questionnaire

- **Satisfaction with the information:** A single question concerning patients' satisfaction with the information they received during their treatment. Numeric scale rating (NRS) scores were used to measure the outcome, ranging from 0 (not satisfied at all) to 10 (extremely satisfied).
- **Level of knowledge:** The patient's perceived level of knowledge about their cycle, administration of hormone injections, side effects of hormone injections, other medications, oocyte retrieval, embryo transfer, and pregnancy test was determined through 7 questions. All questions were scored on a 1 to 5 scale: very knowledgeable, knowledgeable, neutral, little knowledge, and very little knowledge. Sum scores ranging from 7 to 35 were used to measure the outcome.
- **General satisfaction of in vitro fertilization (IVF) intake consultation:** One question assessed patients' overall satisfaction with the IVF intake consultation. NRS scores were used to measure the outcome, ranging from 0 (not satisfied at all) to 10 (extremely satisfied).
- **Ability to understand the information during the IVF intake consultation:** One question addressed patients' ability to understand the information presented during the IVF intake consultation. NRS scores were used to measure the outcome, ranging from 0 (no understanding at all) to 10 (full understanding).
- **Administering hormone injections:** One question evaluated patients' ability to administer the hormone injections at the right time. NRS score was used to measure the outcome, ranging from 0 (not capable at all) to 10 (perfectly capable).
- **Managing side-effects:** One question assessed patients' ability to manage treatment side effects caused by the hormone injections. NRS score was used to measure the outcome, ranging from 0 (not capable at all) to 10 (perfectly capable).
- **Overall quality of the IVF treatment:** The QPP-IVF (Quality from the Patient's Perspective of In Vitro Fertilization) questionnaire [22] assessed 3 dimensions of IVF care: medical-technical conditions (pain, physical care, and waiting time), physical-technical conditions (care room characteristics), and identity-orientated approaches (information during and after treatment, participation, responsibility or continuity, the staffs' respect, and empathy).
- **Health care consumption:** Five questions addressed contacting the hospital in the past 7 days (in addition to planned calls or visits), medication usage, side effects, oocyte retrieval, or other topics. A 0 to 4 score was used to indicate the number of contacts.
- **App usage data:** Continuous logging of all the actions that patients perform in the app, such as opening the app, reading the information, and watching a video.

The study outcomes were measured a total of 4 times during the IVF or ICSI process (Textbox 2). The baseline measurement was taken after patients were enrolled in the study. Follow-up questionnaires were sent to both groups 2 days and 10 days after the IVF intake consultation and 5 days after the oocyte retrieval. Patients were invited to participate in the questionnaire by email. A maximum of 2 email reminders was sent if patients did not respond to the initial invite. Patients had a 7-day window to complete the questionnaires for each measurement. All outcome data were self-reported and collected using an online system. Patients who either missed the baseline measurement or more than 2 follow-up questionnaires were registered as lost to follow-up. These patients were not included in the final data analysis.

Textbox 2. Overview of outcomes assessed throughout the in vitro fertilization (IVF) or intracytoplasmic sperm injection (ICSI) treatment process.

Outcomes assessed during each stage of the IVF or ICSI

- **Baseline:** Satisfaction with the information provided, level of knowledge, and app usage data.
- **2 days after IVF intake:** Satisfaction with the information provided, level of knowledge, general satisfaction with the IVF intake consultation, ability to understand the information during the IVF intake consultation, administering hormone injections, managing side effects, and app usage data.
- **10 days after IVF intake:** Administering hormone injections, managing side effects, app usage data, and health care consumption
- **5 days after oocyte retrieval:** Satisfaction with the information provided, level of knowledge, overall quality of IVF treatment, and app usage data

Sample Size

The sample size calculation was based on the 2016 study, which assessed IVF patients' experiences and satisfaction with patient information [23]. This study revealed an average satisfaction score of 7.29 (SD 2.2) on a 0 to 10 scale. In our study, we expected an average satisfaction of 8.5 (SD 1.5). We performed a power calculation on powerandsamplesize.com using 2-sided equality, α=.05, and β=.90, resulting in 33 patients in each arm. We also added a 10% dropout margin for a total of 36 patients in each arm.

Randomization

Patients were randomized to either the control or intervention group by a computer program. Randomization was performed without block or stratification restrictions. After being allocated to one of the groups, patients received an email that included the link to the baseline questionnaire.

Statistical Methods

For our analysis, we used an intention-to-treat approach, including all randomized patients. Normally distributed continuous variables (eg, satisfaction and level of knowledge) were presented as a mean value with the SD, and they were statistically compared between the groups using independent 2-tailed student *t*-tests. Nonnormally distributed variables were presented as a median value with the IQR. Categorical variables (eg, health care consumption) were presented as sample number and percentage and compared between groups using chi-square tests. Missing data were not replaced in any type of analysis. Patients' level of education was divided into 2 groups for analysis: group 1 (none, elementary school, or secondary or vocational education) and group 2 (higher secondary education, pre-university education, or university education in applied sciences). *P* values of ≤.05 indicated a significant difference, and *P* values between .05 and .10 were indicated a trend. All data were analyzed using SPSS Statistics for Macintosh (version 25.0; IBM).

Results

Study Sample

Between June 2018 and August 2019, a total of 65 patients were willing to participate in the study of which, 2 patients got pregnant before the start of their treatment, 1 patient withdrew due to mental instability, and 8 patients dropped out due to logistical reasons. As a result, a total of 54 patients were randomized into the control and intervention groups.

Of the 54 patients in the study, 4 (7.4%) did not complete the baseline questionnaire, and 2 (3.7%) withdrew from the study for reasons unknown. In total, 28 patients were actively enrolled in the intervention group and 20 patients in the control group. In the intervention group, 25 (89.3%) participants downloaded and used the app (Figure 2). Baseline characteristics of the study population were largely similar between groups (Table 1).

Figure 2. Patient flow diagram. OHSS: ovarian hyperstimulation syndrome.

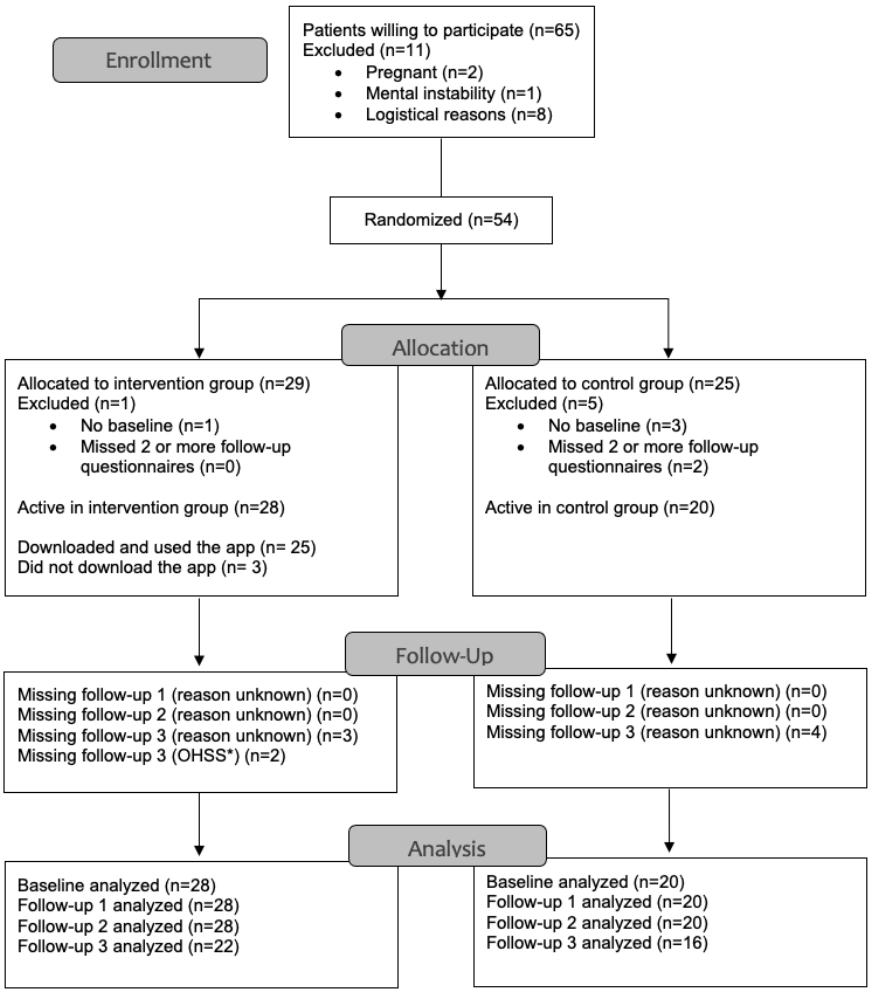


Table 1. Patient characteristics.

Characteristics	Intervention group (n=28)	Control group (n=20)
Age (years), mean (SD)	32.39 (5.14)	32.76 (4.55)
Education, n (%)		
Group 1 (low)	13 (46.4)	10 (50.0)
Group 2 (high)	15 (53.6)	10 (50.0)
Years trying to get pregnant, median (IQR)	2 (1.0-3.0)	2 (1.0-3.0)
Prior IUI[a] treatment at ETZ[b] hospital (yes), n (%)	2 (7.14)	3 (15.0)
Treated before in another hospital (yes), n (%)	4 (14.29)	2 (10.0)

[a]IUI: intrauterine insemination.

[b]ETZ: Elisabeth-TweeSteden Ziekenhuis hospital.

Primary Outcome

Patient Satisfaction With the Information Received During the Treatment

There was no difference between the 2 groups at baseline (intervention group: mean 6.56, SD 1.18 vs control group: mean 6.86, SD 1.18; *P*=.76). However, there was a significant difference in favor of the intervention group 2 days after the IVF intake consultation (intervention group: mean 8.43, SD 1.03 vs control group: mean 7.70, SD 0.66; *P*=.004). At the third and final measurement, 5 days after the oocyte retrieval, the level of satisfaction was equal between groups (intervention group: mean 8.14, SD 1.04 vs control group: mean 8.06, SD 1.44; *P*=.86; Table 2).

Table 2. Patient satisfaction with the information received during the treatment.

Satisfaction with information	Baseline	2 days after IVF[a] intake	5 days after oocyte retrieval
Intervention group, mean (SD), participants	6.56 (1.18) n=28	8.43 (1.03) n=28	8.14 (1.04) n=22
Control group, mean (SD), participants	6.86 (1.18) n=20	7.70 (0.66) n=20	8.06 (1.44) n=16
P value	.76	.004	.86

[a]IVF: in vitro fertilization.

Secondary Outcomes

Level of Knowledge

There was no difference in the level of knowledge between the 2 groups at baseline (intervention group: mean 19.00, SD 3.08 vs control group: mean 17.31, SD 3.20; *P*=.09). However, there was a significant difference in favor of the intervention group 2 days after the IVF intake consultation (intervention group: mean 27.29, SD 2.94 vs control group: mean 23.05, SD 2.76; *P*<.001). At the third and final measurement, 5 days after the oocyte retrieval, the level of knowledge was slightly higher in the intervention group, but this difference was no longer significant (intervention group: mean 27.60, SD 3.48 vs control group: mean 27.13, SD 4.01; *P*=.71; Table 3).

Table 3. Level of knowledge.

Level of knowledge	Baseline	2 days after IVF[a] intake	5 days after oocyte retrieval
Intervention group, mean (SD), participants	19.00 (3.08) n=28	27.29 (2.94) n=28	27.60 (3.48) n=22
Control group, mean (SD), participants	17.31 (3.20) n=20	23.05 (2.76) n=20	27.13 (4.01) n=16
P value	.09	<.001	.71

[a]IVF: in vitro fertilization.

Satisfaction With the IVF Intake Consultation and Ability to Understand the Information

Although patients in the intervention group rated the IVF intake consultation higher than patients in the control group, there was no significant difference in their satisfaction levels (intervention group: mean 9.00, SD 8.61 vs control group: mean 8.60, SD 0.94; *P*=.13). However, patients in the intervention group reported a significantly higher score regarding their ability to understand all the information provided during the consultation (intervention group: mean 8.96, SD 1.14 vs control group: mean 7.95, SD 1.36; *P*=.01).

Administering Hormone Injections and Managing Side Effects

Patients' ability to administer the hormone injections was measured 2 days after the IVF intake and showed no differences between groups (intervention group: mean 8.57, SD 1.10 vs control group: mean 8.15, SD 1.73; *P*=.31). This outcome was measured again 10 days after the IVF intake consultation with

similar results (intervention group: mean 8.74, SD 1.20 vs control group: mean 8.26, SD: 1.94; *P*=.23).

Patients' ability to manage the side effects of the hormone injections was measured 2 days after the IVF intake consultation and showed no differences between groups (intervention group: mean 7.43, SD 1.60 vs control group: mean 7.15, SD 1.47; *P*=.54). It was measured again 10 days after the IVF intake consultation with similar results (intervention group: mean 7.52, SD 1.06 vs control group: mean 7.42, SD 1.43; *P*=.79).

Overall Quality of the IVF Treatment

There was no difference between the groups regarding the perceived overall quality of the IVF treatment (intervention group: mean 51.65, SD 12.73 vs control group: mean 48.59, SD 9.55; *P*=.41).

Health Care Consumption

A trend was observed between the 2 groups concerning health care consumption. Patients in the intervention group contacted the hospital less frequently (intervention group: mean 0.44 contacts per patient, SD 0.85 vs control group: mean 0.84 contacts per patient, SD 0.69; *P*=.09).

App Usage Data

In total, 25 patients in the intervention group used the app 1425 times, an average of 57 times per patient. Patients primarily used a smartphone to access the information (1283/1425, 90%) compared to tablet use (142/1425, 10%). Hormone injection instructions, the side effects of medication, the first day of the IVF cycle, the oocyte retrieval, and usage of the medication capsules after the embryo transfer were consulted most frequently. During the intervention, 26 videos were offered to each patient on average. In total, these videos were viewed 618 times, an average of 24 views per patient. In addition, video-enriched information items about the start of the IVF cycle, medication side effects, oocyte retrieval, and embryo transfer were frequently viewed.

Post-Hoc Power Analysis

Unfortunately, we could not include as many patients as we required based on the initial power calculation we performed. Therefore, to determine the strength of our results, we performed a post-hoc power calculation based on the results of our primary outcome (ie, satisfaction with the information two days after the IVF intake). It showed a power of 81%, indicating that our study was not underpowered.

Discussion

Principal Findings

The results of our study demonstrate the effectiveness of using an app to educate and support patients that undergo IVF treatment. Regarding the primary outcome, patients in the intervention group were more satisfied with the information they received, especially in the first stages of their treatment. Furthermore, the app positively affected patients' knowledge about the different aspects of their IVF treatment and their ability to understand the information during the intake consultation.

To our knowledge, our study is the first to assess the effectiveness of using an app to educate IVF patients through the different stages of their treatment, offering the information promptly by using push notifications. The results on outcomes such as satisfaction and level of knowledge are in line with a 2020 review on using apps to educate patients in this timely manner [19]. Satisfactory patient information and the feeling of a (virtual) continuity of care have been indirectly associated with better individual well-being by reducing treatment concerns and enabling higher treatment tolerability [24]. The importance of information provisioning was also demonstrated by a large European study focusing on patient-centered care in fertility clinics [25]. Being more responsive to patients' needs and expectations can lower the number of discontinued treatments because it reduces the level of emotional distress [26]. However, the results on medication adherence and the management of side effects differ from previous studies, where patients reported anxiety concerning these topics [11,27]. In our study, participants in both groups reported similar positive scores on their ability to manage their medication regimens.

Strengths and Limitations

An important strength of our study is content development, for which we combined multiple insights from specialized fertility physicians and nurses, and embryologists. Another strength is push notifications, allowing the app to reach out to patients when new information was available actively. By delivering the most relevant information in smaller segments, patients can better process and retain the information [19,28]. The app usage statistics demonstrate patients' willingness and need for the app, and that offering complex topics such as the start of the cycle, the oocyte retrieval, side effects of medication, and the embryo transfer through video is highly appreciated.

Limitations

An essential limitation of the study is that we could not include as many patients as required based on the initial power calculation. It was mainly due to staffing problems at the hospital. Nevertheless, a post-hoc power analysis was performed to determine the strength of our results, demonstrating that our study was not underpowered. In addition, we did not involve patient input when deciding which content to offer through the app, including the format and timing of push notifications. This could have contributed to a more personalized experience. Finally, we used several self-reported questionnaires. Although not scientifically validated, we presented the questionnaires to patients prior to the study initiation and used the commonly applied 0 to 10 numeric rating scale mechanism to score the items.

Clinical Implications and Future Research

All patients in the intervention group could download and use the app with no additional instructions besides those provided in the initial email, positively demonstrating the ease and acceptance of the intervention from a patient perspective. It also means that implementing such an app does not require the hospital staff to alter their routines to support patients with the app. In its current form, the app optimizes the IVF patient journey without requiring additional staff resources.

Future development and research should focus on a more personalized version of the app, including a more qualitative approach to better identify patients' needs and new strategies to ensure the information suits their communication styles [29-31]. In addition, adjusting the information based on a patient's anxiety or depression during the different stages of the treatment can enable patients to better cope with the emotional burden of undergoing the treatment and managing related outcomes [32-34].

The app use might be extended to other phases of the IVF treatments or other treatments as well, not only to inform patients about the next step in their treatment but also to actively involve them in their treatment. For example, it would be interesting to see if the patient and health care provider's reported quality of the first consultation would change when patients primarily use the app for educational purposes and in-person clinic visits are spent addressing questions and their priorities. In addition, using the app as a communication platform between patients and health care providers during the entire care journey could further optimize the perceived quality of care. Previous studies focusing on online IVF platforms reported positive outcomes on such ideas, but only if they are strategically implemented in the clinic as integral to the standard of care [35,36].

Conclusions

Our study demonstrates that, in comparison with standard patient education, using an app to provide patients with timely information increases their level of satisfaction. Furthermore, using the app leads to a higher level of knowledge regarding the steps and procedures of IVF treatment. The app's usage statistics demonstrate patients' need for information and their willingness to use an eHealth application as part of their treatment. Future interventions might use a better patient-centered approach, for instance, collecting information about patients' needs and expectations in preparation for their initial consultations or the treatment itself.

Acknowledgments

The authors would like to thank all the staff at the ETZ fertility center for their contribution to the success of this study. Furthermore, they would like to thank the team at Interactive Studios for creating the app. Finally, they would like to thank all patients who were willing to participate in the study.

Authors' Contributions

TT conceived the study and designed the trial with the help of LJ, MK, JAMK, and JS. TT and MK supervised and managed the data collection. TT and LJ provided statistical analysis. JS and MK provided the content used on the patient's timeline. TT and MK drafted the manuscript. All authors contributed to its revision.

Conflicts of Interest

TT, the principal investigator, is one of the cofounders of Interactive Studios, the company that developed the app used in this study. Interactive Studios offered the app free of charge. The co-authors declare that the research was conducted in the absence of any other commercial or financial relationships that could be construed as a potential conflict of interest. Moreover, all authors have completed the ICMJE (International Committee of Medical Journal Editors) uniform disclosure form and declare the following: no support from any organization for the submitted work, no financial relationships with any organizations that might have an interest in the submitted work in the previous 3 years, and no other relationships or activities that could appear to have influenced the submitted work.

Multimedia Appendix 1
Overview and timing of the content in the app.
[PDF File (Adobe PDF File), 54 KB - mhealth_v9i8e28104_app1.pdf]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1169 KB - mhealth_v9i8e28104_app2.pdf]

References

1. Wang J, Sauer MV. In vitro fertilization (IVF): a review of 3 decades of clinical innovation and technological advancement. Therapeutics and Clinical Risk Management 2006 Dec;2(4):355-364. [doi: 10.2147/tcrm.2006.2.4.355]
2. De Geyter C. ART in Europe, 2015: results generated from European registries by ESHRE. Hum Reprod Open 2020;2020(1):001-010. [doi: 10.1093/hropen/hoaa038] [Medline: 32123753]
3. Kissin D. M.., Fertility Treatments in the United States: Improving Access and Outcomes. Obstet Gynecol 2016;128(2):387-390. [doi: 10.1097/aog.0000000000001419]

4. European IVF-monitoring Consortium (EIM) for the European Society of Human Reproduction and Embryology (ESHRE), Wyns C, Bergh C, Calhaz-Jorge C, De Geyter C, Kupka MS, et al. ART in Europe, 2016: results generated from European registries by ESHRE. Hum Reprod Open 2020;2020(3):001-010. [Medline: 32760812]
5. Barrière P, Avril C, Benmahmoud-Zoubir A, Bénard N, Dejager S. Patient perceptions and understanding of treatment instructions for ovarian stimulation during infertility treatment. Reproductive Biomedicine & Society Online 2019 Dec;9:37-47. [doi: 10.1016/j.rbms.2019.08.003] [Medline: 31993512]
6. Mourad S, Hermens RPMG, Cox-Witbraad T, Grol RPTM, Nelen WLDM, Kremer JAM. Information provision in fertility care: a call for improvement. Hum Reprod 2009;24(6):a. [doi: 10.1093/humrep/dep029] [Medline: 19224886]
7. Culley LA, Hudson N, Rapport FL, Katbamna S, Johnson MRD. British South Asian communities and infertility services. Human Fertility 2009 Aug 26;9(1):37-45. [doi: 10.1080/14647270500282644] [Medline: 16581720]
8. Ezabadi Z, Mollaahmadi F, Mohammadi M, Omani Samani R, Vesali S. Identification of Reproductive Education Needs of Infertile Clients Undergoing Assisted Reproduction Treatment Using Assessments of Their Knowledge and Attitude. Int J Fertil Steril 2017;11(1):20-27 [FREE Full text] [doi: 10.22074/ijfs.2016.4728] [Medline: 28367301]
9. Porter M, Bhattacharya S. Helping themselves to get pregnant: a qualitative longitudinal study on the information-seeking behaviour of infertile couples. Human Reproduction 2008 Mar 01;23(3):567-572. [doi: 10.1093/humrep/dem398] [Medline: 18089553]
10. Stewart DE, Rosen B, Irvine J, Ritvo P, Shapiro H, Murphy J, et al. The disconnect: infertility patients' information and the role they wish to play in decision making. Medscape Womens Health 2001 Aug;6(4):1. [Medline: 11547268]
11. Ying L, Wu LH, Loke AY. Gender differences in emotional reactions to in vitro fertilization treatment: a systematic review. J Assist Reprod Genet 2015 Dec 29;33(2):167-179. [doi: 10.1007/s10815-015-0638-4] [Medline: 26712577]
12. Ahmadizadeh S, Bozorgi AS, Kashani L. The role of information therapy in reducing anxiety in patients undergoing fertilisation treatment. Health Info Libr J 2017 Feb 28;34(1):86-91. [doi: 10.1111/hir.12169] [Medline: 28244255]
13. Huisman D, Raymakers X, Hoomans E. Understanding the burden of ovarian stimulation: fertility expert and patient perceptions. Reproductive BioMedicine Online 2009 Jan;19:5-10. [doi: 10.1016/s1472-6483(10)60271-4] [Medline: 19891842]
14. Laflont I, Edelmann RJ. Perceived support and counseling needs in relation to fertilization. Journal of Psychosomatic Obstetrics & Gynecology 2009 Jul 07;15(4):183-188. [doi: 10.3109/01674829409025644] [Medline: 7881503]
15. Redshaw M, Hockley C, Davidson LL. A qualitative study of the experience of treatment for infertility among women who successfully became pregnant. Hum Reprod 2007;22(1):295-304. [doi: 10.1093/humrep/del344] [Medline: 16959809]
16. Schmidt L. Infertile couples' assessment of infertility treatment. Acta Obstet Gynecol Scand 1998;77(6):649-653. [doi: 10.1034/j.1600-0412.1998.770612.x] [Medline: 9688243]
17. Souter VL, Penney G, Hopton JL, Templeton AA. Patient satisfaction with the management of infertility. Human Reproduction 1998 Jul 01;13(7):1831-1836. [doi: 10.1093/humrep/13.7.1831] [Medline: 9740434]
18. Rothwell E, Lamb B, Johnson E, Gurtcheff S, Riches N, Fagan M, et al. Patient perspectives and experiences with fertilization and genetic testing options. Ther Adv Reprod Health 2020;14:2633494119899942 [FREE Full text] [doi: 10.1177/2633494119899942] [Medline: 32518912]
19. Timmers T, Janssen L, Kool RB, Kremer JA. Educating Patients by Providing Timely Information Using Smartphone and Tablet Apps: Systematic Review. J Med Internet Res 2020 Apr 13;22(4):e17342 [FREE Full text] [doi: 10.2196/17342] [Medline: 32281936]
20. 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]
21. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. J Pharmacol Pharmacother 2010 Jul;1(2):100-107 [FREE Full text] [doi: 10.4103/0976-500X.72352] [Medline: 21350618]
22. Holter H, Sandin-Bojö AK, Gejervall A, Wikland M, Wilde-Larsson B, Bergh C. Quality of care in an IVF programme from a patient's perspective: development of a validated instrument. Hum Reprod 2014 Mar;29(3):534-547. [doi: 10.1093/humrep/det421] [Medline: 24287821]
23. Gonen LD. Satisfaction with in vitro fertilization treatment: patients' experiences and professionals' perceptions. Fertil Res Pract 2016;2:6 [FREE Full text] [doi: 10.1186/s40738-016-0019-4] [Medline: 28620533]
24. Gameiro S, Canavarro MC, Boivin J. Patient centred care in infertility health care: direct and indirect associations with wellbeing during treatment. Patient Educ Couns 2013 Dec;93(3):646-654. [doi: 10.1016/j.pec.2013.08.015] [Medline: 24007764]
25. Dancet EAF, D'Hooghe TM, Sermeus W, van Empel I, Strohmer H, Wyns C, et al. Patients from across Europe have similar views on patient-centred care: an international multilingual qualitative study in infertility care. Hum Reprod 2012 Jun;27(6):1702-1711. [doi: 10.1093/humrep/des061] [Medline: 22427309]
26. Brandes M, van der Steen J, Bokdam S, Hamilton C, de Bruin J, Nelen W, et al. When and why do subfertile couples discontinue their fertility care? A longitudinal cohort study in a secondary care subfertility population. Hum Reprod 2009 Dec;24(12):3127-3135. [doi: 10.1093/humrep/dep340] [Medline: 19783833]

27. Sunderam S, Boulet SL, Jamieson DJ, Kissin DM. Effects of patient education on desire for twins and use of elective single embryo transfer procedures during ART treatment: A systematic review. Reprod Biomed Soc Online 2018 Aug;6:102-119 [FREE Full text] [doi: 10.1016/j.rbms.2018.10.017] [Medline: 30761357]
28. Kessels RPC. Patients' memory for medical information. J R Soc Med 2003 May;96(5):219-222 [FREE Full text] [doi: 10.1258/jrsm.96.5.219] [Medline: 12724430]
29. Tuil W, Verhaak C, De Vries Robbé PF, Kremer J. IVF patients show three types of online behaviour. Hum Reprod 2008 Nov;23(11):2501-2505. [doi: 10.1093/humrep/den275] [Medline: 18658162]
30. Van De Belt TH, Hendriks AFW, Aarts JWM, Kremer JAM, Faber MJ, Nelen WLDM. Evaluation of patients' questions to identify gaps in information provision to infertile patients. Hum Fertil (Camb) 2014 Jun;17(2):133-140. [doi: 10.3109/14647273.2014.912762] [Medline: 24869518]
31. Sykes S, Wills J, Frings D, Church S, Wood K. Multidimensional eHealth Literacy for Infertility. Int J Environ Res Public Health 2020 Feb 04;17(3):001-008 [FREE Full text] [doi: 10.3390/ijerph17030966] [Medline: 32033181]
32. De Geyter C, Calhaz-Jorge C, Kupka MS, Wyns C, Mocanu E, Motrenko T, European IVF-monitoring Consortium (EIM) for the European Society of Human ReproductionEmbryology (ESHRE). ART in Europe, 2015: results generated from European registries by ESHRE. Hum Reprod Open 2020;2020(1):hoz038 [FREE Full text] [doi: 10.1093/hropen/hoz038] [Medline: 32123753]
33. Verhaak C, Smeenk JMJ, Evers AWM, Kremer JAM, Kraaimaat FW, Braat DDM. Women's emotional adjustment to IVF: a systematic review of 25 years of research. Hum Reprod Update 2007;13(1):27-36. [doi: 10.1093/humupd/dml040] [Medline: 16940360]
34. Verhaak C, Smeenk JMJ, Nahuis MJ, Kremer JAM, Braat DDM. Long-term psychological adjustment to IVF/ICSI treatment in women. Hum Reprod 2007 Jan;22(1):305-308. [doi: 10.1093/humrep/del355] [Medline: 16973721]
35. Aarts JWM, Faber MJ, Cohlen BJ, Van Oers A, Nelen WLDM, Kremer JAM. Lessons learned from the implementation of an online infertility community into an IVF clinic's daily practice. Hum Fertil (Camb) 2015;18(4):238-247. [doi: 10.3109/14647273.2015.1057901] [Medline: 26167659]
36. Huppelschoten AG, de Bruin JP, Kremer JA. Independent and Web-Based Advice for Infertile Patients Using Fertility Consult: Pilot Study. JMIR Form Res 2019 Jun 04;3(2):13916-13925 [FREE Full text] [doi: 10.2196/13916] [Medline: 31165714]

Abbreviations

CONSORT: consolidated standards for reporting trials
ETZ: Elisabeth-TweeSteden Ziekenhuis hospital
ICSI: intracytoplasmic sperm injection
IVF: in vitro fertilization

Edited by L Buis; submitted 21.02.21; peer-reviewed by B Ma, CI Sartorão Filho; comments to author 22.04.21; revised version received 02.06.21; accepted 02.07.21; published 27.08.21.

Please cite as:
Timmers T, Keijsers M, Kremer JAM, Janssen L, Smeenk J
Supporting Women Undergoing IVF Treatment With Timely Patient Information Through an App: Randomized Controlled Trial
JMIR Mhealth Uhealth 2021;9(8):e28104
URL: https://mhealth.jmir.org/2021/8/e28104
doi:10.2196/28104
PMID:34448725

Original Paper

Methods for Authenticating Participants in Fully Web-Based Mobile App Trials from the iReach Project: Cross-sectional Study

Jodie L Guest[1,2], MPH, PhD; Elizabeth Adam[1], MPH; Iaah L Lucas[1], MPH; Cristian J Chandler[1], MPH, PhD; Rebecca Filipowicz[1], MPH, MS; Nicole Luisi[1], MPH, MS; Laura Gravens[1], MPH, PMP, MS; Kingsley Leung[1], MS; Tanaka Chavanduka[3], MPH; Erin E Bonar[3,4,5], PhD; Jose A Bauermeister[6], MPH, PhD; Rob Stephenson[3,7], MA, MSc, PhD; Patrick S Sullivan[1], DVM, PhD

[1]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, United States

[2]School of Medicine, Emory University, Atlanta, GA, United States

[3]Center for Sexuality and Health Disparities, University of Michigan, Ann Arbor, MI, United States

[4]Department of Psychiatry and Addiction Center, University of Michigan, Ann Arbor, MI, United States

[5]Injury Prevention Center, University of Michigan, Ann Arbor, MI, United States

[6]School of Nursing, University of Pennsylvania, Philadelphia, PA, United States

[7]Department of Systems, Population and Leadership, School of Nursing, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Jodie L Guest, MPH, PhD
Department of Epidemiology
Rollins School of Public Health
Emory University
1518 Clifton Rd Suite 462
Atlanta, GA, 30322
United States
Phone: 1 4046802450
Email: Jodie.Guest@emory.edu

Abstract

Background: Mobile health apps are important interventions that increase the scale and reach of prevention services, including HIV testing and prevention counseling, pre-exposure prophylaxis, condom distribution, and education, of which all are required to decrease HIV incidence rates. The use of these web-based apps as well as fully web-based intervention trials can be challenged by the need to remove fraudulent or duplicate entries and authenticate unique trial participants before randomization to protect the integrity of the sample and trial results. It is critical to ensure that the data collected through this modality are valid and reliable.

Objective: The aim of this study is to discuss the electronic and manual authentication strategies for the iReach randomized controlled trial that were used to monitor and prevent fraudulent enrollment.

Methods: iReach is a randomized controlled trial that focused on same-sex attracted, cisgender males (people assigned male at birth who identify as men) aged 13-18 years in the United States and on enrolling people of color and those in rural communities. The data were evaluated by identifying possible duplications in enrollment, identifying potentially fraudulent or ineligible participants through inconsistencies in the data collected at screening and survey data, and reviewing baseline completion times to avoid enrolling bots and those who did not complete the baseline questionnaire. Electronic systems flagged questionable enrollment. Additional manual reviews included the verification of age, IP addresses, email addresses, social media accounts, and completion times for surveys.

Results: The electronic and manual strategies, including the integration of social media profiles, resulted in the identification and prevention of 624 cases of potential fraudulent, duplicative, or ineligible enrollment. A total of 79% (493/624) of the potentially fraudulent or ineligible cases were identified through electronic strategies, thereby reducing the burden of manual authentication for most cases. A case study with a scenario, resolution, and authentication strategy response was included.

Conclusions: As web-based trials are becoming more common, methods for handling suspicious enrollments that compromise data quality have become increasingly important for inclusion in protocols.

International Registered Report Identifier (IRRID): RR2-10.2196/10174

(JMIR Mhealth Uhealth 2021;9(8):e28232) doi:10.2196/28232

KEYWORDS

HIV; mHealth; recruitment; fraud; adolescent MSM; prevention; MSM; RCT; enrollment; data authentication; data quality; methods; participants

Introduction

Background

Web-based trial recruitment, enrollment, and data collection are increasingly common in research, particularly those focused on the use of mobile health (mHealth) apps. The benefits of using web-based methods include faster and cheaper recruitment, particularly in rural areas [1]. Traditional in-person recruitment strategies are more complicated with adolescent men who have sex with men as they are not able to frequent bars and may not attend gay pride events and other common locations for recruitment in studies of men who have sex with men. This, coupled with the common use of social media apps in this age group, makes web-based recruiting efficient and more generalizable. mHealth apps hold promise to increase the provision of prevention services [2,3] and to reach populations such as adolescent men who have sex with men and rural men who have sex with men who may face interpersonal and structural barriers to seeking in-person prevention services [4-6]. These apps and the use of web-based study methods can be particularly useful for adolescent men who have sex with men residing in rural communities who may not have shared their sexuality with their family and friends and where access to services is challenging. The distribution of these apps through trials is more widespread on the internet, and these apps increase the scale and reach of prevention services, including HIV testing and prevention counseling, pre-exposure prophylaxis, condom distribution, and education, all of which are required to decrease incidence rates [7]. However, web-based trials increase the need for careful scrutiny of forms of fraudulent activity both as an issue of data quality (ie, multiple entries and ineligible participants providing inaccurate age to participate) and as an issue of protection for the adolescent men who have sex with men enrolling in the study. It is critical to ensure that the data collected through this modality are valid and reliable [8].

Authentication for fully web-based studies requires a multimodal approach of electronic and manual verification that may require substantial effort compared with traditional in-person studies [9-11]. There are many examples in the literature on the frequency of fraud in web-based studies. In 2019, Ballard et al [12] categorized 28.7% of their web-based surveys as fraudulent and another 10.1% as potentially fraudulent. In addition, a fully web-based youth-specific HIV study in 2008 identified 675 persons suspected of fraudulent enrollment through multimodal processes [9], and an adolescent men who have sex with men–specific survey published in 2013 found 559 fraudulent cases [13]. Two analyses in 2020 examining issues with web-based recruitment in men who have sex with men concluded that fraud was common, that manual methods work but are resource intensive, and that additional research should be completed to find affordable methods to limit fraudulent enrollment in studies [14,15].

Objectives

Trials of mHealth tools for HIV prevention pose unique challenges, including the authentication of potential participants and the prevention of fraudulent attempts to enroll in studies [16,17]. This study describes the authentication and fraud prevention protocols used in the iReach project, a randomized controlled trial (RCT) of a multilevel life skills intervention that uses mobile apps to reduce vulnerability among men adolescent men who have sex with men [18]. We present the multistep validation process for this web-based adolescent trial, which included electronic programmed comparisons; the use of a manual checklist; and fraud detection methods, including social media. We describe the application of these steps in the trial and provide examples and metrics for the more common types of fraudulent activities, including a brief case scenario for illustration. The strategies used could benefit others who are working on recruitment and enrollment in web-based studies.

Methods

iReach Trial Methods

Methods for conducting the ongoing iReach trial have been described elsewhere [18]. In brief, the trial aimed to explore the efficacy of a multilevel life skills intervention delivered through a web app to 499 adolescents (aged 13-18 years), same-sex attracted, cisgender males (people assigned male at birth who were identified as men) in four US regions, and an additional 101 adolescent men who have sex with men nationally. The participants were a racially and ethnically diverse sample with at least 50% (300/600) identity as people of color or from rural communities. After enrollment, eligible participants assigned to the experimental arm had access to the iReach web app over 12 months of the study. Within the web app, they had access to activity-based life skills modules across 14 key life areas, set goals, monitor progress toward these goals, work on these goals using the peer mentor video chat feature, and access to a locator feature to find community resources that are welcoming to lesbian, gay, bisexual, transgender, queer, questioning, and other sexual and gender minority individuals to support a healthy life and achieve their goals. Participants randomized to the control arm of the trial had access to the locator feature of the intervention due to adolescent men who have sex with men's vulnerability to HIV and sexually transmitted infections. At the end of the 12-month period, participants in the control arm were given full access to the iReach web app for 3 additional months. The primary outcomes of the study were cognitive factors linked to the ability to use HIV prevention and behavioral intentions to use HIV prevention. Participants received a US $30 Amazon

gift card for surveys at baseline and 12-month follow-up; control participants received an additional US $30 gift card for a 15-month follow-up. Participants received a US $25 Amazon gift card for the 3-, 6-, and 9-month follow-up surveys. The University of Pennsylvania Institutional Review Board (IRB) served as the IRB of the record for this study. In accordance with the National Institutes of Health Common Rule, the University of Michigan and Emory University IRBs entered IRB authorization agreements with the University of Pennsylvania IRB. A waiver of parental consent was obtained for minor participants under the age of 18 years.

Recruitment

Recruitment was primarily completed through demographically targeted banner advertisements on social media platforms (ie, Snapchat and Facebook). Additional community engagement using print and media advertisements supplemented the social media advertisements with a link to screening eligibility on the web. All recruitment advertisements asked young men to help test a new health app and included photos of racially and ethnically diverse young men and a link to follow for eligibility.

Interested participants who clicked on the banner ads or accessed the screening survey link were screened for eligibility on the web and completed an electronic informed consent process [19].

Authentication Strategies

Automatic Authentication Strategies

A series of automated authentication strategies identified potentially ineligible participants.

Phone and Contact Information Verification

Potential participants were required to submit their mobile phone number on the screener survey to receive a 3-digit verification code, which they received through SMS text messages. After receiving the 3-digit code, potential participants would then enter the screener survey to validate their mobile phone number for the study.

Participants who failed to input the code during screening were not able to continue the screening survey. When a participant entered an incorrect code, the study team was notified, and the participant was offered assistance in receiving and inputting the code. After verification of the 3-digit code, participants submitted the required information (preferred name, email address) and optional additional contact information (home address, social media handles). These data were part of the manual checklist verification and were used throughout the study for ongoing analysis of enrollment and survey data that would trigger additional reviews.

Those who passed the eligibility criteria, verified their mobile phone number, provided contact information that included their zip code, and consented to the study procedures were routed to the baseline questionnaire. Both automatic and manual verification processes were completed on an ongoing basis for new potential participants three times a week. After full verification, the eligible participants were randomized according to the study protocol.

Questionnaire Data Evaluation

Evaluation Methods and Justification

The screening and baseline questionnaire data were evaluated in the following three ways: (1) by identifying possible duplication of contact information with data from previously registered participants, the team ensured that the participants were new and unique; (2) by identifying inconsistencies in data collected at screening and survey data, the team could identify potentially fraudulent or ineligible participants (those who were outside the eligible age group, who did not report same-sex attraction, who reported being HIV positive at baseline, or did not reside in the targeted recruitment areas were deemed ineligible); and (3) by reviewing baseline completion times, the team avoided enrolling participants or bots who did not complete the baseline questionnaire.

Duplication Checks

Duplication checks were initiated for all participants using a baseline questionnaire record. SAS programs were run to check the newly submitted record against all previous baseline questionnaires to check for duplicates of email addresses, mobile numbers, IP addresses, mailing addresses, social media handles, and preferred names. Potential matches identified by the automated checks were manually reviewed by the study staff. If a potential participant had already been evaluated for study enrollment or had submitted multiple screening surveys with inconsistent information, the participant was not passed to the next step (manual verification). Similarly, the electronic participant management system, Study Management and Retention Toolkit (SMART; developed by Emory University Center for AIDS Research), which was used to track, manage, and contact study participants, searches for exact and partial matches in contact information for each new participant in the SMART system.

Data Comparisons

SAS programs compared the screening survey and baseline questionnaire data for each participant to identify conflicting or inconsistent information between the two surveys using the data elements collected.

Age (in years) was collected in the screening survey, and date of birth (DOB) was collected from the baseline questionnaire. If the age of the participant reported in the screener did not match the age calculated from their reported DOB and baseline questionnaire completion date, these records were flagged for manual review.

IP addresses can identify unique users and their locations. Although IP addresses may be transient, can be duplicated due to institutional IP addresses, or can be changed using proxy servers [11], inconsistencies between the location of the IP address and the self-reported address of participants were considered important indications of potentially fraudulent enrollment. An automated program compared the state of the mailing address submitted by the participant and the state recorded through the IP address. In addition, because eligible participants completing the screening survey were immediately referred to the baseline questionnaire, a second program

compared the IP address locations of screening survey completion and baseline questionnaire completion.

Baseline Questionnaire Completion Time

The baseline questionnaire was designed to take approximately 30 minutes. Baseline questionnaires completed in less than 20 minutes were flagged for manual review.

Each participant's baseline record was assigned a status (complete, partial, or duplicate) and a processing date. Reports of completion scores were generated as participants attempted to enroll in the study; therefore, patterns of multiple fraudulent attempts emerged over time and could be readily identified, and potentially fraudulent participants prevented from enrolling in the study.

Baseline Questionnaire Completion Scores

Participants had the option to skip specific survey questions that they preferred not to answer. To ensure that the enrolled participants were meaningfully engaged in the survey, a random subset of baseline questions that were not impacted by skip patterns was assessed for completion. Participants who completed less than 60% (17/27) of the subset of questions were not referred for enrollment to exclude potential fraud from bots and those who completed research surveys for profit [20]. Similarly, a subsample of the primary outcome questions from the baseline questionnaire was assessed for completion. Participants who completed less than 70% (44/62) of these primary outcome questions were not referred for enrollment. The lack of completion at this early stage may forecast challenges in obtaining complete outcome data within this study, as participants with low completion rates may not understand, recall, or be able to provide judgment on items in the format requested. Participants who surpassed the 60% (17/27) and 70% (44/62) thresholds for the subset of questions and the primary outcome questions, respectively, but completed less than 80% (22/27 and 50/62, respectively) of either question set were flagged for manual review.

Manual Authentication Strategies

Manual Review Process

After the automated authentication checks, a manual review was conducted for participants who flagged for additional review. This manual validation used a checklist (Multimedia Appendix 1), and case report forms were developed to monitor and document the process. A manual review of participants was completed in 1 to 2 business days.

Assessment of Flagged Data

Research staff manually checked all responses of age, IP address comparisons, completion scores, and time stamps, and documented assessments of explainable inconsistencies in the electronic case report forms (eg, a participant identified as 17 years old but added a DOB that was 2 weeks in the future).

Survey Review

If participants had baseline questionnaire completion proportions near the threshold (17/27, 60% for random assortment and 44/62, 70% for primary outcomes), or had multiple flags for review, a manual review of the survey questionnaire was performed to search for patterns of illogical answers (eg, conflicting answers or the same response for all questions or a *Christmas tree* pattern of answering questions, as previously suggested in the literature) [11]. If a pattern was found, it was documented on the manual enrollment checklist, and the participant was not enrolled.

Social Media Review

Although not required for participation, all adolescent men who have sex with men enrolled were asked to submit their social media handles to the study. Of those who provided contact information, 65.18% (863/1324) provided their social media handles. As part of the consent process, participants were asked their permission for the study staff to use these social media platforms to contact the participant or verify their information. The social media review was designed to supplement the enrollment process to verify demographic information (eg, age and gender), the location of the participant, and other helpful information if publicly available on the profile (eg, email address).

SMART Enrollment

The SMART participant management system contains a GPS-specified search engine that allows for the automatic population of zip codes when an address is entered. If the automatically indexed zip code did not match the zip code submitted by the potential participant, this triggered a review for similar addresses and exploration of the address manually. This information was noted in the manual enrollment checklist.

Email Verification

Once all other checks were completed and a determination was made that there was no irregular activity, participants completed an email verification. Participants had 30 days from the issuance of the email verification attempt to respond to the study team. Participants who did not complete this step were not enrolled in the study. Weekly reminders were sent to the participants to increase the likelihood of completing this step.

Ongoing Quality Assurance

Age Verification

iReach collected data every 3 months during the follow-up surveys. The calculated ages from the follow-up surveys were compared with the baseline questionnaire. Variations were manually reviewed, and a determination was made by the study team if participation was discontinued because of concerns about age verification.

Alternate Phone and Email Comparisons

At the time of each follow-up survey, participants were asked to provide additional phone numbers or email addresses as alternatives if the study team could not reach them using their primary contact information. Periodically, the study team evaluated possible matches between each of the alternative contact information provided and the contact information from other participants' main phone numbers and email addresses as a form of potential fraud prevention. If exact or partially matching contact information was found, the two participants were flagged and reviewed to determine whether they were duplicate or dually enrolled.

Other Administrative Study Discontinuations

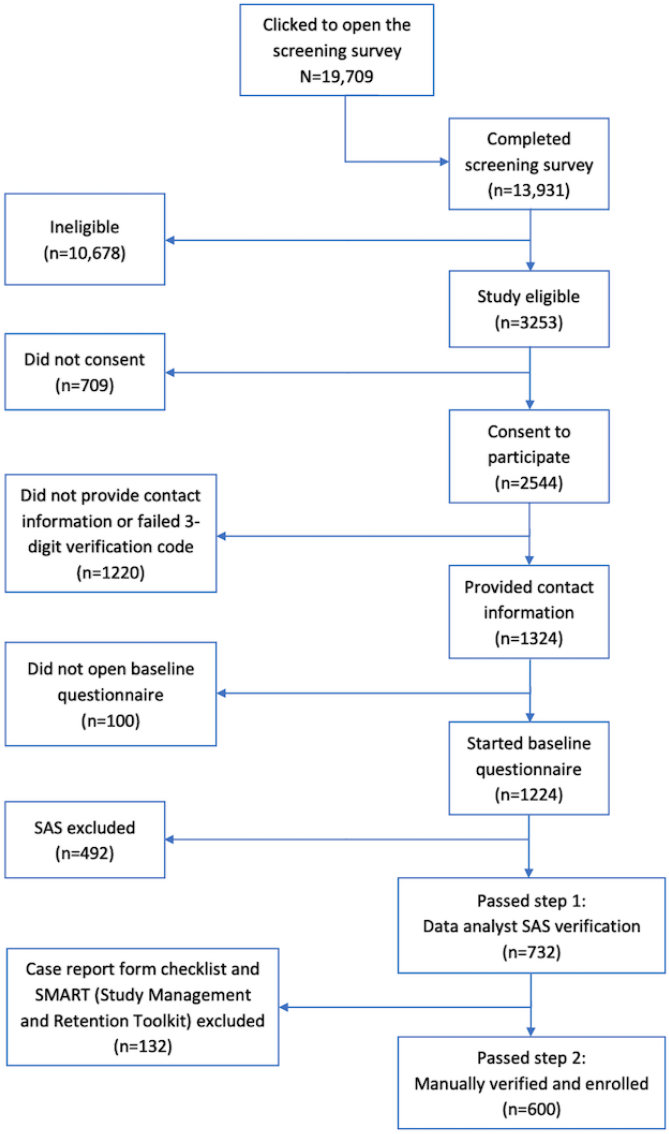
Participants were also removed from the study for any of the following reasons: (1) failure to make contact after multiple engagement attempts, (2) voluntary withdrawal by the participant, (3) errors in the enrollment of the participant that could not be corrected, (4) incarceration of the participant, (5) report of the participant being deceased, or (6) in the event of other unanticipated events that precluded further study participation.

Results

Research Findings

Of the 19,709 visitors to the iReach project screening survey, 13,931 (70.68%) completed the screening survey. Of these, 23.35% (3253/13,931) were eligible for the study. Of the 3253 visitors who were deemed eligible, 2544 (78.2%) consented to participate in the study. After excluding potential participants who did not provide contact information or who failed the 3-digit phone verification, 92.44% (1224/1324) started the baseline questionnaire. Figure 1 shows a flow diagram of enrollment based on the responses of potential participants.

Figure 1. Flow diagram of iReach participant enrollment.



Upon completion of the baseline questionnaire, responses were verified by automatic electronic authentication using the SAS program. This automatic authentication was completed for potential participants in bulk every Monday, Wednesday, and Friday. Table 1 demonstrates that although 1224 participants started the baseline questionnaire, 492 were excluded because of SAS-programmed automatic authentication failures, most of which (n=252) excluded at this stage were excluded because they did not complete 60% (17/27) of the random subset of questions or 70% (44/62) of primary outcome questions. As iReach is a regionally bound RCT for HIV seronegative youth, IP addresses outside the United States were excluded from moving forward (n=8). In addition, several duplications of participants (n=177) who already existed within the study or

the SMART participant management system were excluded. The remaining participants were referred for manual verification.

Table 1. Participant status resulting from automated authentication (N=1224; started baseline).

Status	Participants, n (%)
Failed SAS eligibility recheck[a]	32 (2.61)
Duplication or already enrolled	177 (14.46)
IP address outside the United States	8 (0.65)
Failed completion score requirements[b]	252 (20.59)
Case study participant attempts to enroll	23 (1.88)
Passed, referred to manual authentication	732 (59.8)

[a]Did not meet race criteria: n=30; did not meet HIV status criteria: n=2.

[b]<70% of primary outcome questions or <60% of the random subset of questions.

Tables 2 and 3 show the verification and failure of potential participants referred for manual authentication, respectively. The manual review process was used to exclude 132 individuals (Multimedia Appendix 1). As several areas of data were assessed for irregularities before exclusion, the causes for exclusion in this stage were not mutually exclusive but are noted in Table 3.

Table 2. Participant status resulting from manual authentication (n=732; referred).

Status	Participants, n (%)
Failed, no checklist completed (duplicate)	9 (1.2)
Failed manual checklist and removed	123 (16.8)
Passed and enrolled	600 (81.9)

Table 3. Manual authentication failure reasons for potential participants who were removed from study enrollment (n=123).

Reasons failed[a]	Participants, n (%)
Time stamp fail	33 (26.8)
Age comparison screener and baseline fail	20 (16.3)
Duplicate check fail	28 (22.8)
Suspicious pattern survey response fail	33 (26.8)
Social media check fail (if provided)	20 (16.3)

[a]Failure to move on to enrollment was based on the manual review checklist. The numbers reported are not mutually exclusive.

A total of 132 potential participants failed manual authentication during manual checklist completion. The reasons for failure during manual authentication varied and could overlap if potential participants had multiple issues with the data they provided. The most common reasons for manual authentication failure were a time stamp failure (n=33) or suspicious survey response patterns (n=33). A time stamp failure occurred when potential participants had unusually short or unusually long completion times for their screener survey or baseline questionnaire. A suspicious survey response pattern occurred when the potential participant provided conflicting responses or responded in a pattern (eg, selecting the same response for every answer). Potential participants were also excluded if manual authentication revealed that they did not meet eligibility age requirements for the study, if their social media profile or profiles (if provided) revealed that the potential participant was not who they said they were (ie, did not meet eligibility requirements for age, gender, and location), or if it was discovered that they were duplicates who managed to pass through the automatic authentication process. One specific cluster of 23 potentially fraudulent enrollments was highlighted in the case study provided.

Case Study: The Wisteria Participant

While manually verifying a potential participant for the iReach project, a study team member noted that the zip code provided by the potential participant did not match the SMART-derived zip code (note that the street name has been changed). This flagged an electronic review of the participant data that found the address provided was a partial match with an already-enrolled participant and differed only by house number. In addition, the contact phone number provided by the potential participant was the secondary phone number of the currently enrolled participants with the same street name.

A study team member attempted to contact the potential participant to ask for additional verification to ensure that the potential participant could be authenticated. When the potential participant did not respond after multiple attempts at contact,

they were not enrolled in the study. The previously enrolled participants with a similar address and phone number were sent a message that attempts to duplicate enrollment would be grounds for dismissal from the study.

Twenty-three additional attempts to enroll in the study came from different house numbers on Wisteria Street. Owing to the electronic and manual verification systems in place in the study, multiple attempts for duplication of enrollment were identified and prevented. The enrolled participants were contacted and advised that they were being removed from the study. This example demonstrates the effectiveness of using electronic strategies that include both exact matches and partial matches for addresses to discover potential fraud cases.

Discussion

Principal Findings

This study examines methods for authenticating unique participants for the iReach project, a fully web-based RCT multilevel skills intervention for HIV prevention with adolescents, a population that is easier to enroll on the web. Using a multimodal approach to authentication, 624 potential participants were excluded from enrollment, including those who attempted to enroll more than once. Most participants were excluded by automated data reviews, with a smaller number requiring manual authentication by the study staff.

Over the past 15 years, researchers have acknowledged the threat that potentially fraudulent cases can have on the internal and external validity of their trials [8,9,17]. Several of the authentication checks used in the iReach project mirror those of previous studies, including the electronic verification strategies of IP addresses, the comparison of demographic information at multiple time points (ie, age), and survey time stamp review. Similar to other human verification strategies, manual reviews that look for patterns of illogical responses and requests for the submission of additional proof of eligibility were also included in iReach. The iReach project has also extended previously reported verification processes by improving the methods of possible fraud detection, particularly through the use of electronic authentication strategies. By using automated systems to identify exact and partial matches of demographic information (eg, name and email address) and the SMART participant management system to verify addresses based on GPS locations, many possibly fraudulent cases were eliminated before manual authentication approaches were engaged. Furthermore, the completion thresholds used for all questions and primary outcome questions—60% (17/27) and 70% (44/62), respectively—reduced the number of possibly fraudulent cases that underwent manual authentication. An additional highlight of the iReach project approach was the use of social media for verification. Given that estimates of social media use among youth are as high as 97% [21], using social media data can increase the sensitivity of detecting possible fraudulent enrollments, and some researchers have gone a step further by asking participants to provide a current selfie to match to social media profiles, as described by Bonar et al [22]. Although not a requirement for this study, most participants provided a social media profile. Finally, the use of a single report provided to the study team that detailed possible inaccuracies offered an efficient checklist to ensure a systematic approach to manual authentication.

Strengths and Limitations

It is important to consider how best to characterize the sensitivity and specificity of fraud detection systems. Enrolling fraudulent participants introduces bias into the data, and the detection consumes resources. When working with youth, authentication can limit the potential harm of youth inadvertently interacting with fraudulent accounts or nonminors. However, there is a balance needed in fraud detection, as we strive to include a diversity of participants in the studies. Automated electronic fraud detection methods have the potential to introduce selection bias, as it is plausible that residents of high-density housing developments are at higher risk of being classified as potentially fraudulent based on the similarity of addresses or shared IP addresses than residents of single-family dwellings. This could also be true for participants in college, given that this study included those aged 18 years. A fraud detection system that is only automated might make 1 determination, whereas a fraud detection system that uses both automated and manual verification will be more likely to uncover the reasons for the similarities. Furthermore, 22.8% (28/123) of potential participants who failed manual authentication were discovered to be duplicates, even though they had avoided detection during the initial automatic authentication process, which demonstrates the importance of a combination of both automated and manual authentication for data quality. Potential iReach participants whose data indicated a need for manual review were not automatically or always excluded due to the manual check if staff found explanations for the issue that triggered the manual review. This included participants with long completion times for the baseline questionnaire or those with an IP address that was different from their state address, often due to travel or moving. Although manual reviews certainly take longer and are more time consuming for staff, the targeted use of manual review triggered by the automated review of data resolved issues and inconsistencies by contacting participants. As suggested by Bauermeister et al [23] and Ballard et al [12], this focused manual review allows opportunities to resolve more nuanced situations and improve the specificity of the fraud detection algorithm and keep these participants in the study if the manual review is passed. Whenever possible, study teams strive to strike a balance between automatic and manual authentication strategies to produce the highest quality of research.

Conclusions

Research teams recruiting on the web should be vigilant to maintain scientific rigor in the methods of recruitment and retention. As technology continues to advance, researchers should periodically update methods to ensure the authenticity and uniqueness of participants. Reviews of the design and implementation of electronic and manual strategies for authentication should be performed periodically to ensure that the validity of the study sample is maintained. The careful construction of ways to avert fraud in the design stages can help prepare research teams for unanticipated challenges within this

environment, save time and money with detection efforts, and preserve data quality.

Acknowledgments

This study was supported by a U01 grant from the National Institute of Minority Health and Health Disparities (grant 1U01MD011274-01) and the Centers for AIDS Research at the University of Pennsylvania (grant P30AI045008) and at Emory University (grant P30AI050409).

Conflicts of Interest

None declared.

Multimedia Appendix 1
iReach enrollment checklist.
[DOCX File , 160 KB - mhealth_v9i8e28232_app1.docx]

References

1. Bowen AM, Daniel CM, Williams ML, Baird GL. Identifying multiple submissions in Internet research: preserving data integrity. AIDS Behav 2008 Nov 1;12(6):964-973 [FREE Full text] [doi: 10.1007/s10461-007-9352-2] [Medline: 18240015]
2. Maloney KM, Bratcher A, Wilkerson R, Sullivan PS. Electronic and other new media technology interventions for HIV care and prevention: a systematic review. J Int AIDS Soc 2020 Jan;23(1):e25439 [FREE Full text] [doi: 10.1002/jia2.25439] [Medline: 31909896]
3. Muessig KE, Pike EC, LeGrand S, Hightow-Weidman LB. Mobile phone applications for the care and prevention of HIV and other sexually transmitted diseases: a review. J Med Internet Res 2013 Jan 04;15(1):e1 [FREE Full text] [doi: 10.2196/jmir.2301] [Medline: 23291245]
4. Holloway IW, Rice E, Gibbs J, Winetrobe H, Dunlap S, Rhoades H. Acceptability of smartphone application-based HIV prevention among young men who have sex with men. AIDS Behav 2014 Feb 1;18(2):285-296 [FREE Full text] [doi: 10.1007/s10461-013-0671-1] [Medline: 24292281]
5. Zou H, Fan S. Characteristics of Men Who Have Sex With Men Who Use Smartphone Geosocial Networking Applications and Implications for HIV Interventions: A Systematic Review and Meta-Analysis. Arch Sex Behav 2017 May;46(4):885-894. [doi: 10.1007/s10508-016-0709-3] [Medline: 27043837]
6. Muessig KE, Pike EC, Fowler B, LeGrand S, Parsons JT, Bull SS, et al. Putting prevention in their pockets: developing mobile phone-based HIV interventions for black men who have sex with men. AIDS Patient Care STDS 2013 Apr;27(4):211-222 [FREE Full text] [doi: 10.1089/apc.2012.0404] [Medline: 23565925]
7. Hamilton DT, Goodreau SM, Jenness SM, Sullivan PS, Wang LY, Dunville RL, et al. Potential Impact of HIV Preexposure Prophylaxis Among Black and White Adolescent Sexual Minority Males. Am J Public Health 2018 Nov;108(S4):S284-S291. [doi: 10.2105/ajph.2018.304471]
8. Bauermeister JA, Pingel E, Zimmerman M, Couper M, Carballo-Diéguez A, Strecher VJ. Data Quality in HIV/AIDS Web-Based Surveys. Field Methods 2012 Apr 26;24(3):272-291. [doi: 10.1177/1525822x12443097]
9. Bull SS, Vallejos D, Levine D, Ortiz C. Improving recruitment and retention for an online randomized controlled trial: experience from the Youthnet study. AIDS Care 2008 Sep;20(8):887-893. [doi: 10.1080/09540120701771697] [Medline: 18777217]
10. Chiasson MA, Parsons JT, Tesoriero JM, Carballo-Dieguez A, Hirshfield S, Remien RH. HIV behavioral research online. J Urban Health 2006 Jan 14;83(1):73-85 [FREE Full text] [doi: 10.1007/s11524-005-9008-3] [Medline: 16736356]
11. Kramer J, Rubin A, Coster W, Helmuth E, Hermos J, Rosenbloom D, et al. Strategies to address participant misrepresentation for eligibility in Web-based research. Int J Methods Psychiatr Res 2014 Mar 16;23(1):120-129 [FREE Full text] [doi: 10.1002/mpr.1415] [Medline: 24431134]
12. Ballard AM, Cardwell T, Young AM. Fraud Detection Protocol for Web-Based Research Among Men Who Have Sex With Men: Development and Descriptive Evaluation. JMIR Public Health Surveill 2019 Feb 04;5(1):e12344 [FREE Full text] [doi: 10.2196/12344] [Medline: 30714944]
13. Veinot TC, Meadowbrooke CC, Loveluck J, Hickok A, Bauermeister JA. How "community" matters for how people interact with information: mixed methods study of young men who have sex with other men. J Med Internet Res 2013 Feb 21;15(2):e33 [FREE Full text] [doi: 10.2196/jmir.2370] [Medline: 23428825]
14. Mitchell JW, Chavanduka TM, Sullivan S, Stephenson R. Recommendations from a descriptive evaluation to improve screening procedures for web-based studies with couples: cross-sectional study. JMIR Public Health Surveill 2020 May 12;6(2):e15079-e15626 [FREE Full text] [doi: 10.2196/15079] [Medline: 32396133]

15. Stephenson R, Chavanduka TMD, Sullivan S, Mitchell JW. Correlates of successful enrollment of same-sex male couples into a web-based hiv prevention research study: cross-sectional study. JMIR Public Health Surveill 2020 Jan 9;6(1):e15078 [FREE Full text] [doi: 10.2196/15078] [Medline: 31917373]
16. Pequegnat W, Rosser BRS, Bowen AM, Bull SS, DiClemente RJ, Bockting WO, et al. Conducting Internet-based HIV/STD prevention survey research: considerations in design and evaluation. AIDS Behav 2007 Jul 20;11(4):505-521. [doi: 10.1007/s10461-006-9172-9] [Medline: 17053853]
17. Teitcher JEF, Bockting WO, Bauermeister JA, Hoefer CJ, Miner MH, Klitzman RL. Detecting, preventing, and responding to "fraudsters" in internet research: ethics and tradeoffs. J Law Med Ethics 2015;43(1):116-133 [FREE Full text] [doi: 10.1111/jlme.12200] [Medline: 25846043]
18. Bauermeister J, Sullivan PS, Gravens L, Wolfe J, Countryman K, Smith-Bankhead N, et al. Reducing HIV Vulnerability Through a Multilevel Life Skills Intervention for Adolescent Men (The iREACH Project): Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2018 Jul 10;7(7):e10174 [FREE Full text] [doi: 10.2196/10174] [Medline: 29991470]
19. Hall EW, Sanchez TH, Stein AD, Stephenson R, Zlotorzynska M, Sineath RC, et al. Use of Videos Improves Informed Consent Comprehension in Web-Based Surveys Among Internet-Using Men Who Have Sex With Men: A Randomized Controlled Trial. J Med Internet Res 2017 Mar 06;19(3):e64 [FREE Full text] [doi: 10.2196/jmir.6710] [Medline: 28264794]
20. Pozzar R, Hammer MJ, Underhill-Blazey M, Wright AA, Tulsky JA, Hong F, et al. Threats of Bots and Other Bad Actors to Data Quality Following Research Participant Recruitment Through Social Media: Cross-Sectional Questionnaire. J Med Internet Res 2020 Oct 07;22(10):e23021 [FREE Full text] [doi: 10.2196/23021] [Medline: 33026360]
21. Anderson M, Jiang J. Teens' Social Media Habits and Experiences. Pew Research Center. 2018 Nov 28. URL: https://www.pewresearch.org/internet/2018/11/28/teens-social-media-habits-and-experiences/ [accessed 2019-01-25]
22. Bonar EE, Schneeberger DM, Bourque C, Bauermeister JA, Young SD, Blow FC, et al. Social Media Interventions for Risky Drinking Among Adolescents and Emerging Adults: Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2020 May 13;9(5):e16688 [FREE Full text] [doi: 10.2196/16688] [Medline: 32401225]
23. Bauermeister JA, Zimmerman MA, Johns MM, Glowacki P, Stoddard S, Volz E. Innovative recruitment using online networks: lessons learned from an online study of alcohol and other drug use utilizing a web-based, respondent-driven sampling (webRDS) strategy. J Stud Alcohol Drugs 2012 Sep;73(5):834-838 [FREE Full text] [doi: 10.15288/jsad.2012.73.834] [Medline: 22846248]

Abbreviations

DOB: date of birth
IRB: Institutional Review Board
mHealth: mobile health
RCT: randomized controlled trial
SMART: Study Management and Retention Toolkit

Edited by L Buis; submitted 26.03.21; peer-reviewed by T Winder, AAS Sawitri; comments to author 28.05.21; revised version received 09.06.21; accepted 17.06.21; published 31.08.21.

Please cite as:
Guest JL, Adam E, Lucas IL, Chandler CJ, Filipowicz R, Luisi N, Gravens L, Leung K, Chavanduka T, Bonar EE, Bauermeister JA, Stephenson R, Sullivan PS
Methods for Authenticating Participants in Fully Web-Based Mobile App Trials from the iReach Project: Cross-sectional Study
JMIR Mhealth Uhealth 2021;9(8):e28232
URL: https://mhealth.jmir.org/2021/8/e28232
doi:10.2196/28232
PMID:34463631

Original Paper

Promoting Health via mHealth Applications Using a French Version of the Mobile App Rating Scale: Adaptation and Validation Study

Ina Saliasi[1*], PhD; Prescilla Martinon[1*], MSc; Emily Darlington[1], PhD; Colette Smentek[1], PhD; Delphine Tardivo[2], PhD; Denis Bourgeois[1], PhD; Claude Dussart[1], PhD; Florence Carrouel[1*], PhD; Laurie Fraticelli[1*], PhD

[1]Health, Systemic, Process, Research Unit 4129, University Claude Bernard Lyon 1, University of Lyon, Lyon, France

[2]Laboratory Anthropology, Health Law, and Medical Ethics, UMR 7268, Aix-Marseille University 2, Marseille, France

[*]these authors contributed equally

Corresponding Author:
Ina Saliasi, PhD
Health, Systemic, Process, Research Unit 4129
University Claude Bernard Lyon 1
University of Lyon
11 rue Guillaume Paradin
Lyon, Cedex 08
France
Phone: 33 0478785745
Email: ina.saliasi@univ-lyon.fr

Abstract

Background: In the recent decades, the number of apps promoting health behaviors and health-related strategies and interventions has increased alongside the number of smartphone users. Nevertheless, the validity process for measuring and reporting app quality remains unsatisfactory for health professionals and end users and represents a public health concern. The Mobile Application Rating Scale (MARS) is a tool validated and widely used in the scientific literature to evaluate and compare mHealth app functionalities. However, MARS is not adapted to the French culture nor to the language.

Objective: This study aims to translate, adapt, and validate the equivalent French version of MARS (ie, MARS-F).

Methods: The original MARS was first translated to French by two independent bilingual scientists, and their common version was blind back-translated twice by two native English speakers, culminating in a final well-established MARS-F. Its comprehensibility was then evaluated by 6 individuals (3 researchers and 3 nonacademics), and the final MARS-F version was created. Two bilingual raters independently completed the evaluation of 63 apps using MARS and MARS-F. Interrater reliability was assessed using intraclass correlation coefficients. In addition, internal consistency and validity of both scales were assessed. Mokken scale analysis was used to investigate the scalability of both MARS and MARS-F.

Results: MARS-F had a good alignment with the original MARS, with properties comparable between the two scales. The correlation coefficients (*r*) between the corresponding dimensions of MARS and MARS-F ranged from 0.97 to 0.99. The internal consistencies of the MARS-F dimensions *engagement* (ω=0.79), *functionality* (ω=0.79), *esthetics* (ω=0.78), and *information quality* (ω=0.61) were acceptable and that for the overall MARS score (ω=0.86) was good. Mokken scale analysis revealed a strong scalability for MARS (Loevinger H=0.37) and a good scalability for MARS-F (H=0.35).

Conclusions: MARS-F is a valid tool, and it would serve as a crucial aid for researchers, health care professionals, public health authorities, and interested third parties, to assess the quality of mHealth apps in French-speaking countries.

(JMIR Mhealth Uhealth 2021;9(8):e30480) doi:10.2196/30480

KEYWORDS

mobile health apps; eHealth; Mobile App Rating Scale; MARS; quality assessment tool; rating scale evolution; validation; mHealth; mHealth applications; health applications; mobile health; digital health; digital health tools; application validation

Introduction

In the last few decades, smartphones have radically modified our daily life, as seen by the increasing number of smartphone users worldwide. In parallel to this, an exponential growth of mobile health (mHealth) apps has been observed [1]. Such apps offer an attractive and promising interface for health education and community health promotion [2]. mHealth apps are currently becoming handheld devices that can disseminate a variety of health-promoting knowledge and promote healthy behaviors relating, for example, to dietary habits [3], weight control [4], physical activity [5], addictive behaviors (ie, smoking), and mental health (ie, managing stress and depression) [1]. mHealth apps represent an alternative to or complement face-to-face communication between health care professionals and users of the health care system for primary prevention [6], as well as patients for secondary prevention [7]. They offer an affordable platform that reaches a large audience with possible positive implications for public health, especially health promotion and prevention strategies [1].

Before the deployment of an app on the web, the app store reviews it as well as its updates, in order to determine whether it is reliable, performs as expected, respects user privacy, and is free of objectionable content such as offensive language or nudity. However, the review by the developer is not comprehensive enough to enable end users, health professionals, and researchers to identify and evaluate the quality of mHealth apps [8,9]. The most common way to select an mHealth app that is currently available on the app market is by using publicly available information, and by considering easily available attributes such as title, price, star ratings, reviews, or downloads, instead of validated scientific content [10]. To date, certification and trust labels for mobile apps are not widely endorsed [11].

Few mHealth apps available on the market have undergone a thorough validation process based on high-level evidence that can be a potential problem for the safety of end users [9]. In order to evaluate the validity and functionality of mHealth apps objectively, several standardized scales have been developed for health care professionals [12]. The Mobile Application Rating Scale (MARS) was developed by Stoyanov et al [8] in the English language, and, to date, it is considered the reference scale for health care professionals in the scientific literature. The Italian, Spanish, German, and Arabic versions of MARS have already been produced and validated [2,13-15]. The 23-item scale assesses the quality of health-related apps through four objective dimensions relating to the quality of the mHealth app (engagement, functionality, esthetics, and information) and one subjective dimension (subjective app quality and perceived impact).

The aim of this study is to develop and validate a French version of the Mobile App Rating Scale (MARS-F) as a multidimensional measure for trialing, classifying, and rating the quality of mHealth apps.

Methods

Study Design

The validation of this study followed and applied a well-established process of cross-cultural adaptation [16], translation and back-translation, review, piloting, and psychometric evaluation.

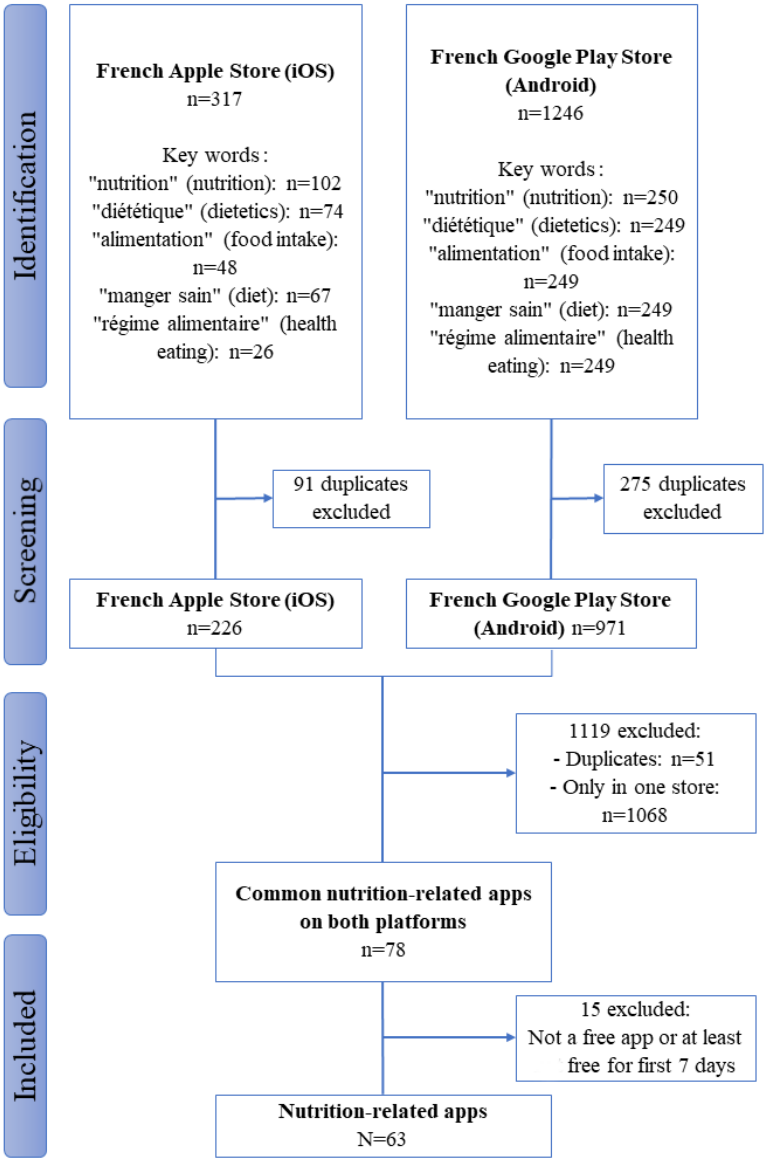
Cultural Adaptation and Translation

First, the translation of MARS from English to French was conducted by two independent bilingual scientists (IS and LF). Following the review, discussion, and comparison of their two forward translations, they agreed upon a common pilot version of MARS-F. Second, this common pilot version was blind back-translated by two bilingual native English speakers with different educational backgrounds—a researcher in public health and educational sciences (ED) and a nonacademic professional (ADB). Third, the two bilingual scientists (IS and LF) compared the back-translated version with the original English version. After mutual discussion, they agreed upon the final French version of the scale (MARS-F). Finally, 6 other people (3 researchers and 3 nonacademic professionals) evaluated the comprehensibility of this finalized French version. Their comments were considered, and the final MARS-F version was thus created (Multimedia Appendix 1).

Selection of Apps

The inclusion process consisted of three different phases: searching, screening, and determining the eligibility criteria of nutrition health-related apps. The search for apps was conducted from March 10, 2021, to March 17, 2021, on the French Apple Store (iOS) and Google Play Store (Android). No truncation or use of logic operators (AND, OR, and NOT) was possible while searching in the Google Play Store and iOS Store. Hence, in order to select the nutrition health-related apps, the following search terms were used separately: “nutrition” (nutrition), “diététique” (dietetics), “alimentation” (food intake), “régime alimentaire” (diet), and “manger sain” (healthy eating). Apps were included if they were available free of charge or at least free of charge during 7 days from both the iOS Store and Google Play Store. Duplicate copies of apps between the two stores were excluded, resulting in a total of 63 apps (Figure 1).

Figure 1. Flowchart of the app selection process.



Raters' Training

To complete the evaluation process of apps, we used the rating methodology previously described by Stoyanov et al [8]. We made a video with an introduction of the French MARS scale, and an exercise on how to rate a nutrition mobile app (available on request to the corresponding author). Two individuals with a master's degree in medical sciences (FC and PM), and who were fluent in both French and English, were instructed on how to serve as raters by watching the video. With a view to ensure that the raters were sufficiently trained, they were asked to download and evaluate 10 apps that were randomly selected from those meeting our inclusion criteria using MARS and MARS-F. Each app rater tested each app for at least 15 minutes before they carried out their evaluation. Raters then compared their individual rating scores for each app. When their individual rating scores varied by at least 2 points, they discussed their findings until they aligned their rating approaches and agreed on the score.

Data Analysis

Intraclass Correlation

The two raters completed the evaluation of the remaining 53 apps independently. The intraclass correlation coefficients (ICCs) were calculated to measure the interrater reliability of the items, the subscales, and total MARS scores with absolute agreement between the raters. An ICC of <0.50 was interpreted as poor; 0.51-0.75, as moderate; 0.76-0.89, as good; and >0.90, as excellent correlation [17]. We excluded item 19 due to missing values.

Internal Consistency

The internal consistency of MARS-F and its subscales were also assessed as a measure of scale reliability, as reported in the original MARS study. We used the omega coefficient instead of the Cronbach alpha coefficient, as it is commonly used to assess reliability as described in the literature. The omega coefficient provides justifiably higher estimates of reliability than the Cronbach alpha coefficient [18]. The robust procedure introduced by Zhang and Yuan was used to estimate omega

values, with the objective to obtain a closer estimate to the population value without being overwhelmingly affected by a few overbearing observations [19]. Reliability was assessed as follows: ω<0.50 was interpreted as unacceptable internal consistency; ω=0.51-0.59, as poor consistency; ω=0.60-0.69, as questionable consistency; ω=0.70-0.79, as acceptable consistency; ω=0.80-0.89, as good consistency; and ω>0.90, as excellent consistency.

Validity

To establish an indicator of validity, we investigated the subscale correlations between MARS-F and its original English version. In addition, we calculated the overall correlation between the total MARS score and total MARS-F score. The correlation coefficient ranges between –1 and 1. The closer the coefficient is to 1, the stronger the positive linear relationship between the variables. The closer the coefficient is to –1, the stronger the negative linear relationship between the variables. Mean comparisons were also performed between the corresponding dimensions of MARS and MARS-F, and *P* values were adjusted for multiple testing according to Holmes' method [20]. For all dimensions compared, we considered a *P* value <.05 as statistically significant.

Mokken Scale Analysis

Mokken scale analysis (MSA) is a technique used for scaling test and questionnaire data closely. This technique is related to the nonparametric item response theory [21,22]. The latent monotonicity and nonintersection are two necessary preconditions to use the MSA. Loevinger H is the key parameter of this scale. For item *i*, the scaling parameter is H_i, and H_k is the scaling parameter for the overall scalability of all items in the scale k. H_i indicates the strength of the relationship between a latent variable (eg, app quality) and item *i*. The SE values of the scalability coefficients of the item pairs were also calculated. A scale is considered weak if H is <0.4, moderate if H is ≥0.4 but <0.49, and strong if H is >0.5 [21,23]. MSA was conducted for both MARS and MARS-F to assess the scalability of the mean scores. As recommended by van der Ark, the reliability of the scales was additionally assessed using the Molenaar-Sijtsma (MS) method [24], λ-2, and latent class reliability coefficient (LCRC) [14,25].

Statistical Analysis

R software (version 4.0.5; R Foundation for Statistical Computing) was used for all analyses. The correlations, ICC, and MSA were conducted using the R packages psych (function corr.test) (version 1.8.12), coefficient alpha (function omega) (version 0.5) and mokken (function coefH) (version 1.8.12). The two preconditions of latent monotonicity and nonintersection were tested using the functions check.monotonicity and check.restscore from the package mokken. The statistics related to the reliability of the scales were provided using the function check.reliability.

Results

Calibration During Raters' Training

Among the 10 common apps, the mean scores of the dimensions *engagement* (t_{10}=–0.76, *P*=.44), *functionality* (t_{53}=–0.11, *P*=.90), *esthetics* (t_{53}=0.22, *P*=.82), and *information quality* (t_{33}=–0.35, *P*=.72) were equivalent in both versions. The internal consistencies of MARS (ω=0.86, 95% CI 0.56-0.96) and MARS-F (ω=0.78, 95% CI 0.32-0.94) were good and acceptable, respectively, and the MSA revealed strong scalability (H=0.47; SE=0.07).

Descriptive Data and Mean Comparisons

The ICCs for MARS (0.88, 95% CI 0.79-0.93) and MARS-F (0.89, 95% CI 0.8-0.93) were high. The mean and SD scores of the items in MARS and MARS-F are presented in Table 1. The mean scores of the dimensions *engagement* (t_{53}=–0.34, *P*=.72), *functionality* (t_{53}=–0.47, *P*=.63), *esthetics* (t_{53}=.09, *P*=.92), and *information quality* (t_{33}=0, *P*>.99) were equivalent between MARS and MARS-F.

Table 1. Summary of item and scale scores for the original version of the Mobile App Rating Scale (MARS) and the French version of MARS (MARS-F) (n=53 apps).

Dimension	Score, mean (SD)	
	MARS	MARS-F
Engagement	2.83 (0.89)	2.85 (0.88)
Item 1	2.67 (0.81)	2.63 (0.82)
Item 2	2.97 (0.74)	2.90 (0.78)
Item 3	2.65 (0.94)	2.66 (0.92)
Item 4	2.60 (0.99)	2.84 (0.95)
Item 5	3.24 (0.80)	3.20 (0.79)
Functionality	4.32 (0.72)	4.35 (0.72)
Item 6	4.09 (0.87)	4.11 (0.85)
Item 7	4.25 (0.63)	4.29 (0.62)
Item 8	4.30 (0.59)	4.31 (0.64)
Item 9	4.64 (0.64)	4.67 (0.63)
Esthetics	3.33 (0.79)	3.32 (0.82)
Item 10	3.73 (0.64)	3.75 (0.62)
Item 11	3.26 (0.77)	3.23 (0.83)
Item 12	3.00 (0.79)	3.00 (0.80)
Information quality	3.25 (1.08)	3.25 (1.08)
Item 13	3.96 (0.39)	3.96 (0.39)
Item 14	3.78 (0.70)	3.75 (0.71)
Item 15	3.50 (0.75)	3.49 (0.75)
Item 16	3.21 (0.80)	3.22 (0.76)
Item 17	3.42 (0.98)	3.43 (0.99)
Item 18	1.64 (0.90)	1.66 (0.90)
Item 19	N/A[a]	N/A[a]
Overall mean	3.43 (0.43)	3.44 (0.43)

[a]N/A: this item on information quality could not be rated because it was nonapplicable.

Internal Consistency

The internal consistency of MARS and MARS-F and their subscales is presented in Table 2. The internal consistency of the MARS dimension *engagement* (ω=0.82, 95% CI 0.79-0.87) was good. The internal consistencies of the dimensions *functionality* (ω=0.80, 95% CI 0.74-0.85) and *esthetics* (ω=0.79, 95% CI 0.73-0.88) were acceptable and that for *information quality* (ω=0.64, 95% CI 0.49-0.70) indicated questionable consistency. The internal consistency of the overall MARS score was good (ω=0.87, 95% CI 0.83-0.91).

For MARS-F, the internal consistencies were acceptable for the dimensions *engagement* (ω=0.79, 95% CI 0.72-0.83), *functionality* (ω=0.79, 95% CI 0.73-0.85), *esthetics* (ω=0.78, 95% CI 0.71-0.82), and *information quality* (ω=0.61, 95% CI 0.53-0.65). The internal consistency of the overall MARS score was good (ω=0.86, 95% CI 0.85-0.90).

Table 2. Internal consistency per dimension for the original version of the Mobile App Rating Scale (MARS) and the French version of MARS (MARS-F).

Dimension	Internal consistency, ω (95% CI)	
	MARS	MARS-F
Engagement	0.82 (0.79-0.87)	0.79 (0.72-0.83)
Functionality	0.80 (0.74-0.85)	0.79 (0.73-0.85)
Esthetics	0.79 (0.73-0.88)	0.78 (0.71-0.82)
Information quality	0.64 (0.49-0.70)	0.61 (0.53-0.65)
Overall mean	0.87 (0.83-0.91)	0.86 (0.85-0.90)

Validity

The correlation coefficients between the corresponding dimensions of MARS and MARS-F ranged from 0.97 to 0.99. *P* values were adjusted for multiple testing according to Holmes' method (Table 3). Correlations between the respective items are presented in Multimedia Appendix 2.

Table 3. Correlation between the English and French versions of the Mobile App Rating Scale.

Dimension	Engagement FR	Functionality FR	Esthetics FR	Information quality FR
Engagement ENG				
r	0.98	0.30	0.57	0.63
P value	<.001	.02	<.001	<.001
Functionality ENG				
r	0.24	0.98	0.49	0.28
P value	.03	<.001	<.001	.02
Esthetics ENG				
r	0.52	0.50	0.99	0.52
P value	<.001	<.001	<.001	<.001
Information quality ENG				
r	0.66	0.34	0.56	0.97
P value	<.001	.01	<.001	<.001

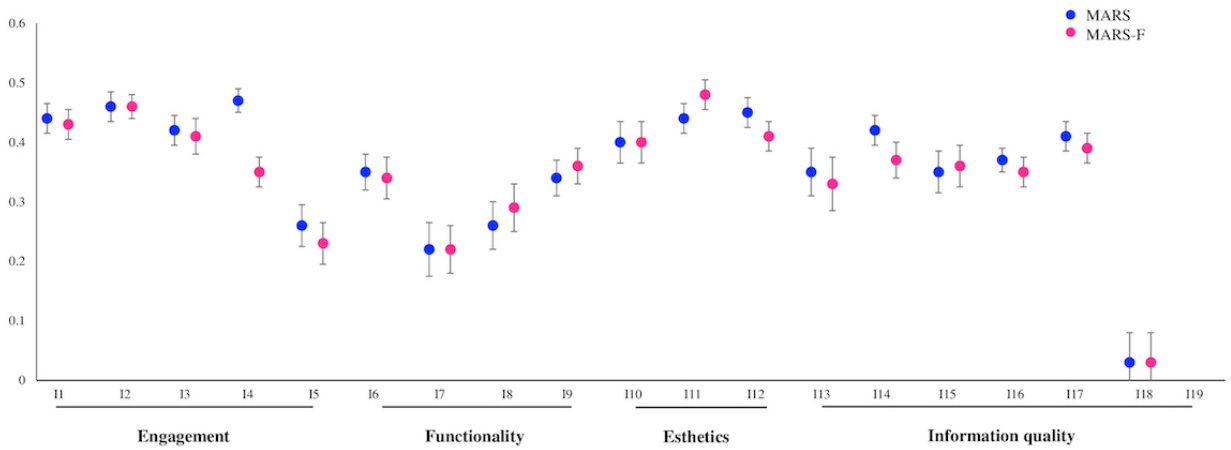
MSA Results

MSA results for both versions of the scale (ie, MARS and MARS-F) are summarized in the Table 4. MSA for MARS revealed strong scalability (H=0.37; SE=0.03). There was no violation of monotonicity because the item step response functions were nondecreasing functions; likewise, there was no violation of nonintersection because the item step response functions do not intersect. The internal consistency of this scale was acceptable (MS=0.88; λ-2=0.88; LCRC=0.89). MSA for MARS-F revealed good scalability (H=0.35; SE=0.03). The internal consistency of this scale was acceptable (MS=0.88; λ-2=0.89; LCRC=0.90). The scalability results of MARS and MARS-F are presented in Figure 2.

Table 4. Mokken scale analysis for the original version of the Mobile App Rating Scale (MARS) and the French version of MARS (MARS-F).

	MARS	MARS-F
Loevinger H coefficient	0.37	0.35
Standard errors of the scalability coefficients of the item pairs	0.03	0.03
Molenaar-Sijtsma coefficient	0.88	0.88
Lambda-2	0.88	0.89
Latent class reliability coefficient	0.89	0.90

Figure 2. Loevinger (H_k) coefficients (overall scalability of all items in the scale) for the Mobile App Rating Scale (MARS) and the French version of the Mobile App Rating Scale (MARS-F) depending on various dimensions.



Discussion

Principal Results

This study aimed to develop and evaluate MARS-F to enable French health care professionals to assess the quality of mHealth apps. To our knowledge, this is the first cultural adaptation, translation, and validity evaluation of the original MARS in French.

Nutrition-related apps were identified using well-defined and selected search terms in both two app stores (Google Play Store and Apple Store). This was done to avoid methodological challenges such as ranking algorithms or irrelevant results because the indexing of apps is usually determined by a developer who is most interested in promoting the app.

With a view to provide a comparable interpretation of statistical indicators, the methodology was chosen to be similar to previous adaptations of the scale [2,8,14,26]. In addition, 63 apps were included, which is higher than the minimum sample size of 41 apps required to confirm that interrater reliability lies within 0.15 of a sample observation of 0.80, with 87% assurance [26]. We used the same strategy that led to the Italian version of MARS except that the team included apps by searching and screening across three app stores (Google Play, Apple, and Windows Stores). As per the validation of the German version of MARS, each search term was provided separately, as no truncation or use of logic operators (AND, OR, and NOT) was possible in the Google Play and Apple Store. In our study, two raters downloaded and then evaluated 10 apps that were randomly selected for training and piloting purposes as in the initial English version of MARS against 5 apps in the Italian [2], Spanish [13], and German development versions [14].

The interrater reliability of MARS and MARS-F were aligned, with overlapping CI values. The ICCs for MARS-F were also comparable to the Arabic (0.84, 95% CI 0.82-0.85) [15] and German versions of MARS (0.83, 95% CI 0.82-0.85) [14], and they were slightly lower than the Italian version of MARS (0.96, CI 0.90-0.91) [2] (Multimedia Appendix 3).

The internal consistency of the overall MARS score was good and that of MARS-F was acceptable for the dimensions *engagement*, *functionality*, *esthetics*, and *information quality*. The internal consistency of the German version of MARS was good for *engagement* and excellent for *functionality* and *esthetics*. On the other hand, the internal consistency of *information quality* was acceptable. For the Arabic version of MARS, the internal consistency was good for *engagement* and *esthetics*, good for *information quality*, and acceptable for *functionality* [15]. All Cronbach alpha coefficients were judged to be at least acceptable for the Italian version of MARS [2], and these values were high for the Spanish [13] version of MARS.

MSA results for MARS-F revealed a good scalability (H=0.35, SE=0.03), and the use of total MARS-F score was found to be appropriate. Additionally, we obtained a high correspondence between MARS-F and the original MARS [8], which demonstrates proven validity.

The same methodology was used for the validation of the German (apps targeting anxiety), Italian (primary prevention), Spanish (health and fitness apps), and Arabic versions (health and fitness apps). Our results were consistent with the findings of the research teams that developed and validated the Italian, Spanish, German, and Arabic versions of the MARS [2,14,15] (Multimedia Appendix 3).

Limitations

The first possible limitation could be that the validation of MARS-F is based on the evaluation of nutrition-related apps, whereas MARS is applicable to mHealth apps. The second limitation could be attributed to the fact that MARS-F was elaborated by native speakers living in France. French speakers can have diverse cultures according to their country. Therefore, further adaptation could be required. The third limitation concerns item 19 on information quality. This item could not be rated because raters choose the response option “non applicable,” which allows raters to skip an item if the app does not contain any health-related information (eg, nutrition apps in this study). The same item was also excluded from all calculations in the Italian version of MARS because of lack of

ratings [2]. This item evaluates the evidence-based literature relating to the nutrition app assessed, and it is worth noting that many apps have not yet been scientifically evaluated.

Future Perspectives

With 300 billion French-speakers worldwide [27], the translation of MARS could be of special interest. Owing to its wide use in the assessment of mHealth apps in the scientific literature, we chose to translate MARS into French to provide a reliable and understandable tool for health professionals to get an evidence-based sense of the quality and reliability of chosen mHealth apps. Other rating scales such as App Quality Evaluation (AQEL) [28], ENLIGHT [29], and the app evaluation model from the American Psychiatric Association [30] could also represent relevant tools to evaluate mHealth apps for further investigations. All these scales were created for the evaluation of mHealth apps, except AQEL that specifically evaluates nutrition-related apps [28]. Several studies have demonstrated that nutrition is one of the key factors in oral and general health [31]. It would be interesting to translate this scale into French and to evaluate the nutrition-related apps included in our study.

Alongside the assessment process of mHealth apps, the patient's involvement in such processes should also be considered. The user version of the MARS (uMARS) [32] should be translated and evaluated for reliability and validity. Mobile technology represents an innovative opportunity to assist end users in improving their management of their chronic conditions. Such in-the-pocket devices could be adapted to the specific needs of populations. As an example, mHealth apps could be used for young people's transition to adult care services [33], to support active adults [34], or to promote healthy aging [35]. mHealth apps are valuable for the primary and secondary prevention of chronic diseases, especially for controlling individual risk factors and preventing the snowball effect of chronic diseases with aging [31].

Conclusions

To conclude, MARS-F would be a crucial aid for researchers, health care professionals, public health authorities, and interested third parties, to assess the quality of mHealth apps in French-speaking countries. In addition, French app developers could use this French version as a tool to evaluate and improve the quality of their apps prior to market launch. MARS-F is an important cornerstone to app quality assessment with the purpose to identify reliable and valid apps for the benefit of end users.

Acknowledgments

We would like to thank Catherine Billard, Adeline Darlington-Bernard, Géraldine Seveste for their contribution in the translation process.

Conflicts of Interest

None declared.

Multimedia Appendix 1
French version of the Mobile Application Rating Scale (MARS-F).
[PDF File (Adobe PDF File), 190 KB - mhealth_v9i8e30480_app1.pdf]

Multimedia Appendix 2
Correlation coefficients between the respective items of the Mobile App Rating Scale (MARS) and the French version of the Mobile App Rating Scale (MARS-F).
[XLSX File (Microsoft Excel File), 12 KB - mhealth_v9i8e30480_app2.xlsx]

Multimedia Appendix 3
Comparison of internal consistencies and the Mokken Scale Analysis between Mobile App Rating Scale in English and other available translations of the scale.
[PDF File (Adobe PDF File), 57 KB - mhealth_v9i8e30480_app3.pdf]

References

1. 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]
2. Domnich A, Arata L, Amicizia D, Signori A, Patrick B, Stoyanov S, et al. Development and validation of the Italian version of the Mobile Application Rating Scale and its generalisability to apps targeting primary prevention. BMC Med Inform Decis Mak 2016;16:83 [FREE Full text] [doi: 10.1186/s12911-016-0323-2] [Medline: 27387434]
3. Rabbi M, Pfammatter A, Zhang M, Spring B, Choudhury T. Automated personalized feedback for physical activity and dietary behavior change with mobile phones: a randomized controlled trial on adults. JMIR Mhealth Uhealth 2015;3(2):e42 [FREE Full text] [doi: 10.2196/mhealth.4160] [Medline: 25977197]

4. Laing BY, Mangione CM, Tseng C, Leng M, Vaisberg E, Mahida M, et al. Effectiveness of a smartphone application for weight loss compared with usual care in overweight primary care patients: a randomized, controlled trial. Ann Intern Med 2014 Nov 18;161(10 Suppl):S5-12. [doi: 10.7326/M13-3005] [Medline: 25402403]
5. Bort-Roig J, Gilson ND, Puig-Ribera A, Contreras RS, Trost SG. Measuring and influencing physical activity with smartphone technology: a systematic review. Sports Med 2014 May;44(5):671-686. [doi: 10.1007/s40279-014-0142-5] [Medline: 24497157]
6. Kampmeijer R, Pavlova M, Tambor M, Golinowska S, Groot W. The use of e-health and m-health tools in health promotion and primary prevention among older adults: a systematic literature review. BMC Health Serv Res 2016 Sep 05;16 Suppl 5:290 [FREE Full text] [doi: 10.1186/s12913-016-1522-3] [Medline: 27608677]
7. Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. P T 2014 May;39(5):356-364 [FREE Full text] [Medline: 24883008]
8. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
9. Llorens-Vernet P, Miró J. Standards for mobile health-related apps: systematic review and development of a guide. JMIR Mhealth Uhealth 2020 Mar 03;8(3):e13057 [FREE Full text] [doi: 10.2196/13057] [Medline: 32130169]
10. Huang H, Bashir M. Users' adoption of mental health apps: examining the impact of information cues. JMIR Mhealth Uhealth 2017 Jun 28;5(6):e83 [FREE Full text] [doi: 10.2196/mhealth.6827] [Medline: 28659256]
11. Singh K, Drouin K, Newmark LP, Lee J, Faxvaag A, Rozenblum R, et al. Many mobile health apps target high-need, high-cost populations, but gaps remain. Health Aff (Millwood) 2016 Dec 01;35(12):2310-2318. [doi: 10.1377/hlthaff.2016.0578] [Medline: 27920321]
12. Azad-Khaneghah P, Neubauer N, Miguel Cruz A, Liu L. Mobile health app usability and quality rating scales: a systematic review. Disabil Rehabil Assist Technol 2020 Jan 08:1-10. [doi: 10.1080/17483107.2019.1701103] [Medline: 31910687]
13. Martin Payo R, Fernandez Álvarez MM, Blanco Díaz M, Cuesta Izquierdo M, Stoyanov SR, Llaneza Suárez E. Spanish adaptation and validation of the Mobile Application Rating Scale questionnaire. Int J Med Inform 2019 Sep;129:95-99. [doi: 10.1016/j.ijmedinf.2019.06.005] [Medline: 31445295]
14. Messner E, Terhorst Y, Barke A, Baumeister H, Stoyanov S, Hides L, et al. The German version of the Mobile App Rating Scale (MARS-G): development and validation study. JMIR Mhealth Uhealth 2020 Mar 27;8(3):e14479 [FREE Full text] [doi: 10.2196/14479] [Medline: 32217504]
15. Bardus M, Awada N, Ghandour LA, Fares E, Gherbal T, Al-Zanati T, et al. The Arabic version of the Mobile App Rating Scale: development and validation study. JMIR Mhealth Uhealth 2020 Mar 03;8(3):e16956 [FREE Full text] [doi: 10.2196/16956] [Medline: 32130183]
16. Herdman M, Fox-Rushby J, Badia X. A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. Qual Life Res 1998 May;7(4):323-335. [doi: 10.1023/a:1024985930536] [Medline: 9610216]
17. Portney L, Watkins M. Foundations of Clinical Research: Applications to Practice. United States: F.A. Davis Company; 2009.
18. McNeish D. Thanks coefficient alpha, we'll take it from here. Psychol Methods 2018 Sep;23(3):412-433. [doi: 10.1037/met0000144] [Medline: 28557467]
19. Zhang Z, Yuan K. Robust coefficients alpha and omega and confidence intervals with outlying observations and missing data: methods and software. Educ Psychol Meas 2016 Jun;76(3):387-411 [FREE Full text] [doi: 10.1177/0013164415594658] [Medline: 29795870]
20. Holm S. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 1979;6(2):65-70 [FREE Full text]
21. Ark LAVD. Mokken Scale Analysis in R. J Stat Soft 2007;20(11):1-19. [doi: 10.18637/jss.v020.i11]
22. Sijtsma K, van der Ark LA. A tutorial on how to do a Mokken scale analysis on your test and questionnaire data. Br J Math Stat Psychol 2017 Feb;70(1):137-158. [doi: 10.1111/bmsp.12078] [Medline: 27958642]
23. Ark LAVD. New developments in Mokken Scale Analysis in R. J Stat Soft 2012;48(5):1-27 [FREE Full text] [doi: 10.18637/jss.v048.i05]
24. Molenaar I, Sijtsma K. Mokken's approach to reliability estimation extended to multicategory items. Kwantitatieve Methoden: Nieuwsbrief voor Toegepaste Statistiek en Operationele Research? 1988;9(28):115-126.
25. van der Ark LA, van der Palm DW, Sijtsma K. A latent class approach to estimating test-score reliability. Applied Psychological Measurement 2011 Mar 09;35(5):380-392. [doi: 10.1177/0146621610392911]
26. Zou GY. Sample size formulas for estimating intraclass correlation coefficients with precision and assurance. Stat Med 2012 Dec 20;31(29):3972-3981. [doi: 10.1002/sim.5466] [Medline: 22764084]
27. Qui parle français dans le monde – Organisation internationale de la Francophonie. Webpage in French. Langue française et diversité linguistique. URL: http://observatoire.francophonie.org/qui-parle-francais-dans-le-monde/ [accessed 2021-05-03]
28. DiFilippo KN, Huang W, Chapman-Novakofski KM. A new tool for nutrition App Quality Evaluation (AQEL): development, validation, and reliability testing. JMIR Mhealth Uhealth 2017 Oct 27;5(10):e163 [FREE Full text] [doi: 10.2196/mhealth.7441] [Medline: 29079554]

29. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: a comprehensive quality and therapeutic potential evaluation tool for mobile and web-based eHealth interventions. J Med Internet Res 2017 Mar 21;19(3):e82 [FREE Full text] [doi: 10.2196/jmir.7270] [Medline: 28325712]
30. The App Evaluation Model. American Psychiatric Association. URL: https://www.psychiatry.org/psychiatrists/practice/mental-health-apps/the-app-evaluation-model [accessed 2021-05-03]
31. Martinon P, Fraticelli L, Giboreau A, Dussart C, Bourgeois D, Carrouel F. Nutrition as a key modifiable factor for periodontitis and main chronic diseases. J Clin Med 2021 Jan 07;10(2):197 [FREE Full text] [doi: 10.3390/jcm10020197] [Medline: 33430519]
32. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
33. Virella Pérez YI, Medlow S, Ho J, Steinbeck K. Mobile and web-based apps that support self-management and transition in young people with chronic illness: systematic review. J Med Internet Res 2019 Nov 20;21(11):e13579 [FREE Full text] [doi: 10.2196/13579] [Medline: 31746773]
34. Mascarenhas MN, Chan JM, Vittinghoff E, Van Blarigan EL, Hecht F. Increasing physical activity in mothers using video exercise groups and exercise mobile apps: randomized controlled trial. J Med Internet Res 2018 May 18;20(5):e179 [FREE Full text] [doi: 10.2196/jmir.9310] [Medline: 29776899]
35. Chen YR, Schulz PJ. The effect of information communication technology interventions on reducing social isolation in the elderly: a systematic review. J Med Internet Res 2016 Jan 28;18(1):e18 [FREE Full text] [doi: 10.2196/jmir.4596] [Medline: 26822073]

Abbreviations

AQEL: app quality evaluation
ICC: intraclass correlation coefficients
LCRC: latent class reliability coefficient
MARS: Mobile Application Rating Scale
MARS-F: Mobile Application Rating Scale–French
MSA: Mokken Scale Analysis
MS: Molenaar-Sijtsma
uMARS: user version of the Mobile Application Rating Scale

Edited by L Buis; submitted 17.05.21; peer-reviewed by P Phantumvanit, A Kanoute; comments to author 22.06.21; revised version received 25.06.21; accepted 06.07.21; published 31.08.21.

Please cite as:
Saliasi I, Martinon P, Darlington E, Smentek C, Tardivo D, Bourgeois D, Dussart C, Carrouel F, Fraticelli L
Promoting Health via mHealth Applications Using a French Version of the Mobile App Rating Scale: Adaptation and Validation Study
JMIR Mhealth Uhealth 2021;9(8):e30480
URL: https://mhealth.jmir.org/2021/8/e30480
doi:10.2196/30480
PMID:34463623

Original Paper

Enabling Wearable Pulse Transit Time-Based Blood Pressure Estimation for Medically Underserved Areas and Health Equity: Comprehensive Evaluation Study

Venu Ganti[1], BS; Andrew M Carek[1], PhD; Hewon Jung[1], MS; Adith V Srivatsa[2]; Deborah Cherry[3], BBA; Levather Neicey Johnson[3], BSN, RN; Omer T Inan[1,2], PhD

[1]School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, United States

[2]The Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, United States

[3]VSNS, Inc, Atlanta, GA, United States

Corresponding Author:
Venu Ganti, BS
School of Electrical and Computer Engineering
Georgia Institute of Technology
85 5th St NW
Atlanta, GA, 30308
United States
Phone: 1 2406434250
Email: vganti6@gatech.edu

Abstract

Background: Noninvasive and cuffless approaches to monitor blood pressure (BP), in light of their convenience and accuracy, have paved the way toward remote screening and management of hypertension. However, existing noninvasive methodologies, which operate on mechanical, electrical, and optical sensing modalities, have not been thoroughly evaluated in demographically and racially diverse populations. Thus, the potential accuracy of these technologies in populations where they could have the greatest impact has not been sufficiently addressed. This presents challenges in clinical translation due to concerns about perpetuating existing health disparities.

Objective: In this paper, we aim to present findings on the feasibility of a cuffless, wrist-worn, pulse transit time (PTT)–based device for monitoring BP in a diverse population.

Methods: We recruited a diverse population through a collaborative effort with a nonprofit organization working with medically underserved areas in Georgia. We used our custom, multimodal, wrist-worn device to measure the PTT through seismocardiography, as the proximal timing reference, and photoplethysmography, as the distal timing reference. In addition, we created a novel data-driven beat-selection algorithm to reduce noise and improve the robustness of the method. We compared the wearable PTT measurements with those from a finger-cuff continuous BP device over the course of several perturbations used to modulate BP.

Results: Our PTT-based wrist-worn device accurately monitored diastolic blood pressure (DBP) and mean arterial pressure (MAP) in a diverse population (N=44 participants) with a mean absolute difference of 2.90 mm Hg and 3.39 mm Hg for DBP and MAP, respectively, after calibration. Meanwhile, the mean absolute difference of our systolic BP estimation was 5.36 mm Hg, a grade B classification based on the Institute for Electronics and Electrical Engineers standard. We have further demonstrated the ability of our device to capture the commonly observed demographic differences in underlying arterial stiffness.

Conclusions: Accurate DBP and MAP estimation, along with grade B systolic BP estimation, using a convenient wearable device can empower users and facilitate remote BP monitoring in medically underserved areas, thus providing widespread hypertension screening and management for health equity.

(JMIR Mhealth Uhealth 2021;9(8):e27466) doi:10.2196/27466

KEYWORDS

wearable sensing; pulse transit time; cuffless blood pressure; noninvasive blood pressure estimation; health equity; mobile phone

Introduction

Background

Current clinical practice regarding hypertension management and control hinges on the century-old approach of obtaining infrequent cuff-based measurements of blood pressure (BP) in clinical settings. This paradigm of the measurement being anchored to the clinical setting and requiring persons to proactively visit a medical professional to determine their hypertensive status is costly—due to the time and money spent [1]—and considered ineffective—due to the infrequency and error (ie, white coat hypertension) of office BP measurements [2,3]. Hence, we observed remarkable disparities in hypertension detection, treatment, and control across socioeconomic status and race, with populations lacking access to regular office visits and care, having nearly half the awareness of their existing hypertensive status, and enduring up to triple the rates of subsequent cardiac events [4,5]. Technologies that enable frequent, reliable, and accurate measurements of BP in ambulatory settings promise to reduce the global burden of hypertension and offer an opportunity to advance health equity [6]. Leveraging the ubiquity of smartphones and digital health technologies equipped with highly sensitive, miniaturized sensors is essential for the remote monitoring of BP [7].

Existing wearable devices that incorporate noninvasive BP methodologies offer an affordable and efficient means of tracking out-of-office BP [8]. Unfortunately, they commonly use uncomfortable techniques (ie, oscillometry and tonometry) that demand imparting forces on blood vessels to achieve accurate measurements [9-11]. These inconveniences fail to empower users to take control of their health, posing a significant challenge to the widespread routine monitoring of BP. Instead, strategies that compute the pulse transit time (PTT), a measure of arterial stiffness, present a convenient alternative for BP estimation [12].

The PTT, the time the pressure wave propagates along the length of the arterial tree, is a cuffless surrogate for BP and can be acquired noninvasively [12]. In practice, the acquisition of noninvasive PTT requires a combination of sensors—typically an accelerometer, force sensor, light-emitting diode (LED) and photodiode, electrode, or ultrasonic transceiver—placed proximally and distally along the arterial tree and computed from fiducial points in the captured seismocardiogram (SCG), ballistocardiogram, photoplethysmogram (PPG), impedance cardiogram, impedance plethysmogram, or arterial blood pressure (ABP) waveform [13,14]. Despite their inherent convenience, these sensing modalities are naturally of concern when used in populations with intrinsic mechanical, optical, and electrical barriers, stemming from higher melanin levels and body fat percentages.

To the best of our knowledge, noninvasive PTT-based BP estimation has yet to be examined in a diverse population—a gap in our scientific understanding that presents a formidable obstacle to its adoption. Specifically, some medically underserved areas (MUAs), which stand to benefit the most from remote monitoring [15], have a large number of Black and Latino individuals with higher melanin content and obesity rates compared with White individuals [16]. Recent notable data from the Centers for Disease Control and Prevention further stress these concerns by exposing that non-Hispanic Black individuals not only have significantly *higher hypertension prevalence* than non-Hispanic White individuals but also witness significantly *lower hypertension control* rates [17,18]. Affordable remote monitoring options have the responsibility to combat not only social determinants of health, such as access to health care and income, but also in turn the existing health disparities that are byproducts of them. As a result, there exists a glaring hole in PTT-based BP monitoring—that this technology has yet to be tested on the population for whom it may be the most valuable, and until now, its continuing practice will only exacerbate existing health disparities.

Objectives

In our previous work, we designed a wearable, multimodal, wrist-worn PTT monitoring device (SeismoWatch) and validated it in both controlled lab [19] and unsupervised home [20] settings, primarily on young, healthy persons with lighter skin. In this paper, we expand upon our previous work with a community-engaged research strategy that leverages expertise from a nonprofit organization serving MUAs in Georgia and evaluated our device in a more diverse population. We present our device's ability to accurately estimate BP in this diverse population and capture significant demographic-level differences in underlying arterial stiffness that coincide with observations from existing literature, through the calibration coefficients used in our BP estimation model. This work represents the first time that a noninvasive, cuffless, PTT-based wearable device has been extensively evaluated in a community-based diverse population as a potentially reliable and convenient monitoring option toward, ultimately, the remote screening and management of hypertension for health equity.

Methods

Study Protocol

A comprehensive breakdown of the demographics of the study population is presented in Table 1. This study was conducted under a protocol approved by the Georgia Institute of Technology institutional review board (protocol number H19251). The study was separated into two different populations (N=44 participants) referred to throughout this work as follows: (1) a young and healthy homogeneous population (first cohort=26 participants) and (2) an older, entirely Black, higher BMI, metropolitan population (second cohort=18 participants) recruited later through the help of our community outreach partners—a nonprofit organization serving medically underserved persons in the state of Georgia. For the first cohort, 26 (19 males and 7 females) young and healthy volunteers (mean age 26.7 years, SD 3.7; mean weight 73.8 kg, SD 14.1; height 173.9 cm, SD 9.6; and mean BMI 24.2 kg/m^2, SD 3.2) with no previous history of cardiovascular disease were recruited, and written informed consent was obtained. For the second cohort, 18 (6 males and 12 females) Black participants (mean age 44.1 years, SD 11.7; mean weight 94.4 kg, SD 18.0; mean height 169.6 cm, SD 11.5; and mean BMI 33.2 kg/m^2, SD 7.6) with no previous history of cardiovascular disease other than

hypertension were recruited from the Atlanta metropolitan area, written informed consent was obtained, and further demographic information was collected post hoc with verbal consent. Both hypertensive status and the use of regular prescription medications were self-reported.

Table 1. Participant demographics and cardiovascular parameters for study participants (grouped by cohort; N=44).

Demographics and cardiovascular parameters[a]	Homogenous data set (first cohort; n=26; participant 1-26)	Community outreach (metropolitan Atlanta) data set (second cohort; n=18; participant 27-44)	*P* value
Age (years), mean (SD)	26.7 (3.7)	44.1 (11.7)	<.001
Sex, n (%)			
Male	19 (73)	6 (33)	N/A[b]
Female	7 (26)	12 (67)	N/A
Height (cm), mean (SD)	173.9 (9.6)	169.6 (11.5)	.19
Weight (kg), mean (SD)	73.8 (14.1)	94.4 (18.0)	<.001
BMI[c] (kg/m^2), mean (SD)	24.2 (3.2)	33.2 (7.6)	<.001
Obesity class, n (%)			N/A
I	1[d] (4)	2[e] (11)	
II	N/A	3[f] (17)	
III	N/A	4[g] (22)	
Race, n (%)			N/A
Black	1[h] (4)	18 (100)	
Other race	25 (96)	N/A	
Hypertensive status, n (%)			N/A
Normotensive	26 (100)	15 (83)	
Hypertensive	N/A	2[i] (11)	
Hypotensive	N/A	1[j] (6)	
Current medications, n (%)			N/A
Hydrochlorothiazide (1×day)	N/A	2[k] (11)	
Lisinopril (1×day)	N/A	1[l] (6)	
Iron supplement	N/A	1[m] (6)	

[a]Statistical significance between groups in values, where applicable, was computed using an unpaired two-tailed t test.

[b]N/A: not applicable.

[c]Obesity classified using the BMI per the guidelines from the National Heart, Lung, and Blood Institute of the National Institutes of Health [21] (I: BMI=30-34.9; II: BMI=35-39.9; III: BMI ≥40).

[d]Participant 23.

[e]Participants 30 and 43.

[f]Participants 38, 40, and 42.

[g]Participants 34, 36, 37, and 41.

[h]Participant 5.

[i]Participants 29 and 37.

[j]Participant 33.

[k]Participants 29 and 37.

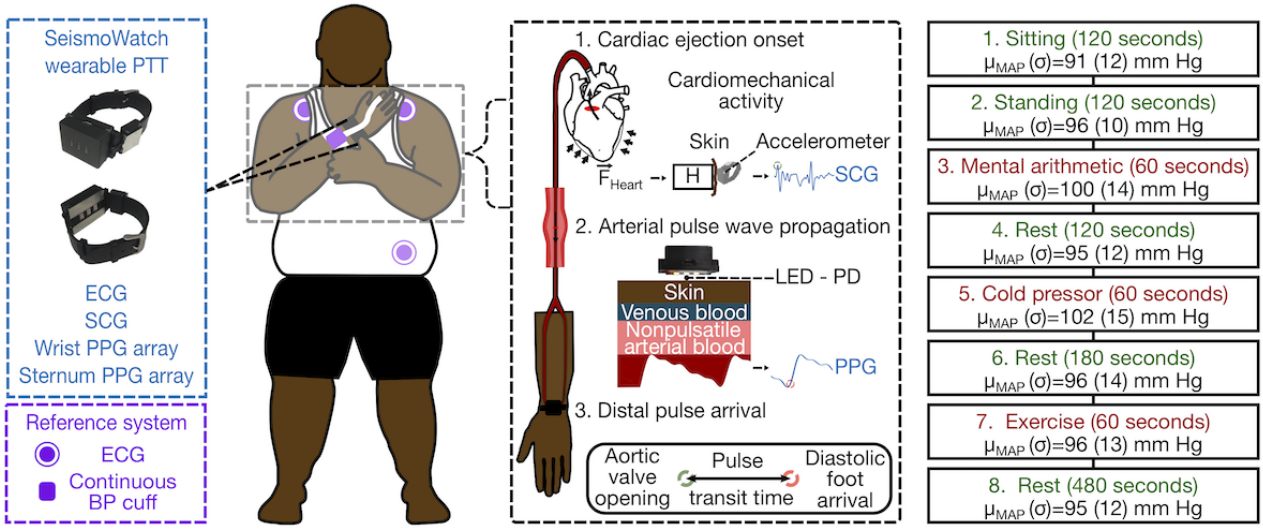
[l]Participant 29.

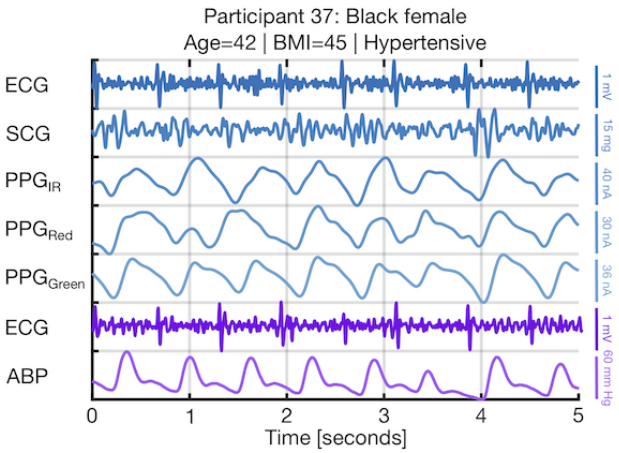
[m]Participant 33.

The concept of the study design is shown in Figure 1. Although not explicitly shown, two versions of the SeismoWatch were used in this study: a previous version of the hardware with comparable sensors was used in the young, homogeneous population (ie, the first cohort), before being adapted for a more robust, portable, and multimodal wearable device used in the metropolitan Atlanta population (ie, the second cohort). Specifically, the data from these cohorts were collected during

two intervals, between which the hardware was revised to incorporate multiwavelength PPGs before investigating the performance of the sensing modality in the underrepresented population. This was essential to assess the efficacy of shorter-wavelength LEDs (ie, those with shallower skin penetration depths) in a Black population. However, in both the correlations in Figure 2 and calibration coefficient comparisons in Figure 3, only the results derived from the infrared (IR) PPGs, available to both devices, were computed and shown. The other key sensing components and reference system components were essentially identical: (1) the first version of the device used an analog version of the accelerometer (ADXL354, Analog Devices) to acquire the SCG, whereas the second version simply used the digital version of the same sensor (ADXL355, Analog Devices) to reduce size and (2) the finger-cuff continuous BP reference system (ccNexfin, Edwards Lifesciences) along with the data acquisition module (MPU150, Biopac Systems) were identical in both studies.

Figure 1. Concept overview and study design. Sensor information and placement locations for wearable system (blue) and reference system (purple). Noninvasive pulse transit time (PTT) measurement concept overview using seismocardiogram (SCG) and photoplethysmogram (PPG) sensors. Study protocol tasks in chronological order with duration and mean (SD) of mean arterial pressure (MAP) values for each task. Sample filtered signals from the participant with the lowest signal-to-noise ratio (SNR) signals (n=37): a hypertensive, high BMI, older Black female. In order from top to bottom: electrocardiogram (ECG), SCG, infrared PPG, red PPG, green PPG signals measured from the wearable system (blue) and the synchronized ECG, and arterial blood pressure (ABP) signals measured by the reference system (purple). Systolic blood pressure (SBP; top) and diastolic blood pressure (DBP; bottom) plotted across the full protocol for participant 37, with rest periods (green) and perturbations used to modulate BP (red) highlighted in chronological order, and the location where the reference finger-cuff continuous blood pressure (BP) system was paused during the exercise indicated. ABP: arterial blood pressure; BP: blood pressure; DBP: diastolic blood pressure; ECG: electrocardiogram; LED: light-emitting diode; PD: photodiode; PPG: photoplethysmogram; PTT: pulse transit time; SBP: systolic blood pressure; SCG: seismocardiogram.





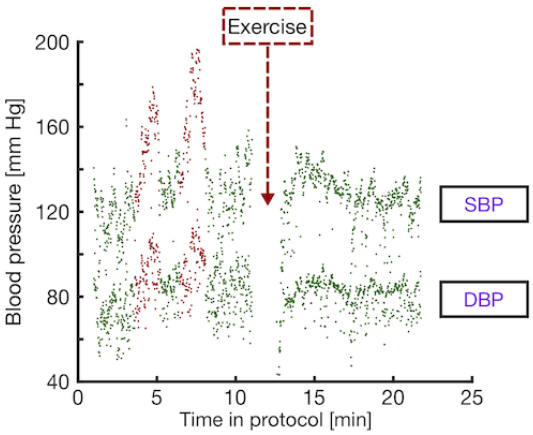


Figure 2. Wearable pulse transit time (PTT)–based blood pressure (BP) estimation results. Correlation and Bland-Altman plots between PTT-estimated BP and the finger-cuff continuous BP for mean arterial pressure, diastolic blood pressure, and systolic blood pressure estimation. The root mean squared error and the mean absolute difference for each correlation are shown. DBP: diastolic blood pressure; MAD: mean absolute difference; MAP: mean arterial pressure; PTT: pulse transit time; RMSE: root mean square error; SBP: systolic blood pressure.

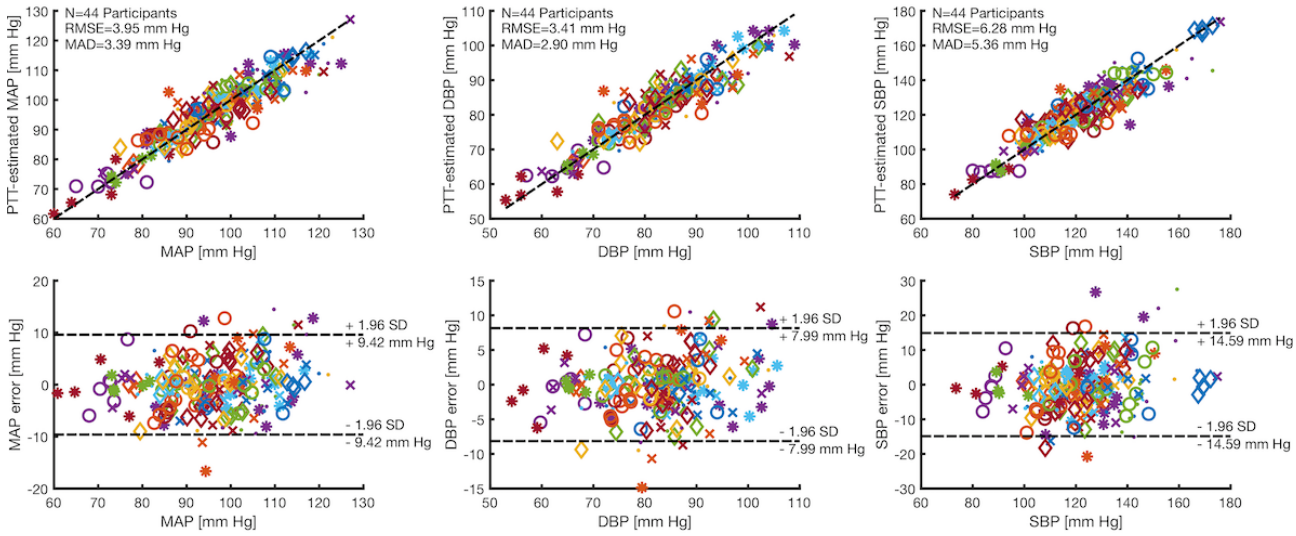
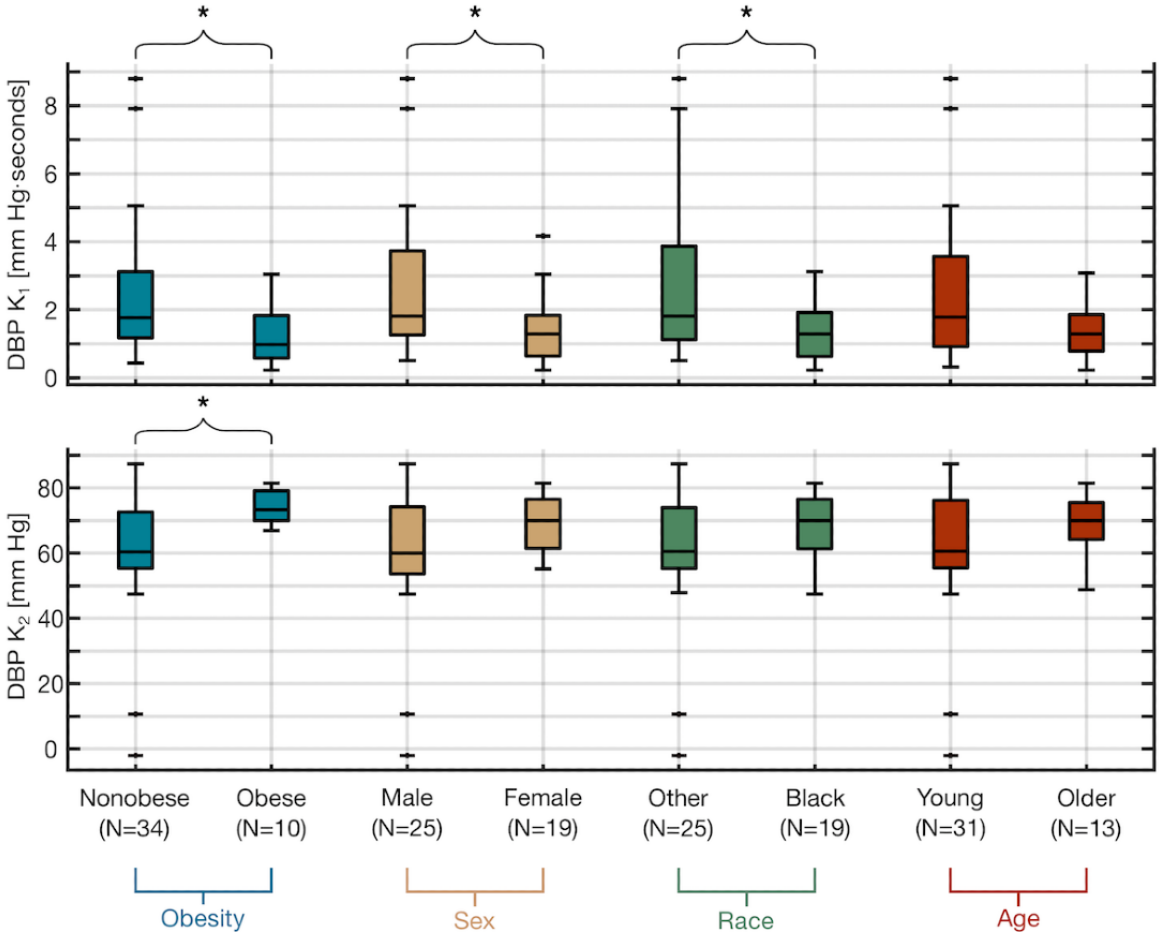


Figure 3. Participant-specific diastolic blood pressure (DBP) calibration coefficients are significantly different in demographics with typical disparities in arterial stiffness. Boxplots showing the statistically significant (*P<.05; Mann-Whitney U) difference in the DBP K1 and K2 calibration coefficients between participants who are nonobese and obese. Boxplots showing the statistically significant (*P<.05; Mann-Whitney U) difference in the DBP K1 calibration coefficients between male and female participants. Boxplots showing the statistically significant (*P<.05; Mann-Whitney U) difference in the DBP K1 calibration coefficients between participants of other race and Black participants. Boxplots showing the difference in the DBP K1 and K2 calibration coefficients between young and older participants. DBP: diastolic blood pressure.



To acquire a timing reference for the start of a cardiac cycle, while serving as a reference for alignment to the wearable system signals, a wireless electrocardiogram (ECG) module (BN-EL50, Biopac Systems) was attached to the participant in a three-lead configuration with Ag/AgCl gel electrodes as shown in Figure 1. As depicted in Figure 1, a finger-cuff BP sensor

based on the volume-clamp methodology (ccNexfin, Edwards Lifesciences) [22,23] was placed on the same hand as the watch, acquiring a reference measurement of continuous beat-by-beat BP. Although volume-clamping continuous BP devices are not the clinical gold standard for ABP measurements, an arterial line was not feasible due to invasiveness, and a sphygmomanometer was not used because of the need for a trained professional and lack of beat-by-beat BP data. Similarly, semiautomated BP cuffs were not used as they hinge on following strict guidelines to obtain an accurate reading, such as being seated and resting the arm at heart level, which were impossible to satisfy simultaneously while acquiring watch measurements, given the need for the contralateral hand to touch the ECG electrode to activate the PTT mode [20]. In addition, it was recently demonstrated that a volume-clamping–based system had comparable accuracy with noninvasive oscillometric BP cuffs [24]. All reference system sensors were sampled at 1 kHz and interfaced to a computer using a data acquisition system (MPU150, Biopac Systems) and its corresponding software (Acqknowledge, Biopac Systems). The reference system files were saved to a desktop computer for postprocessing.

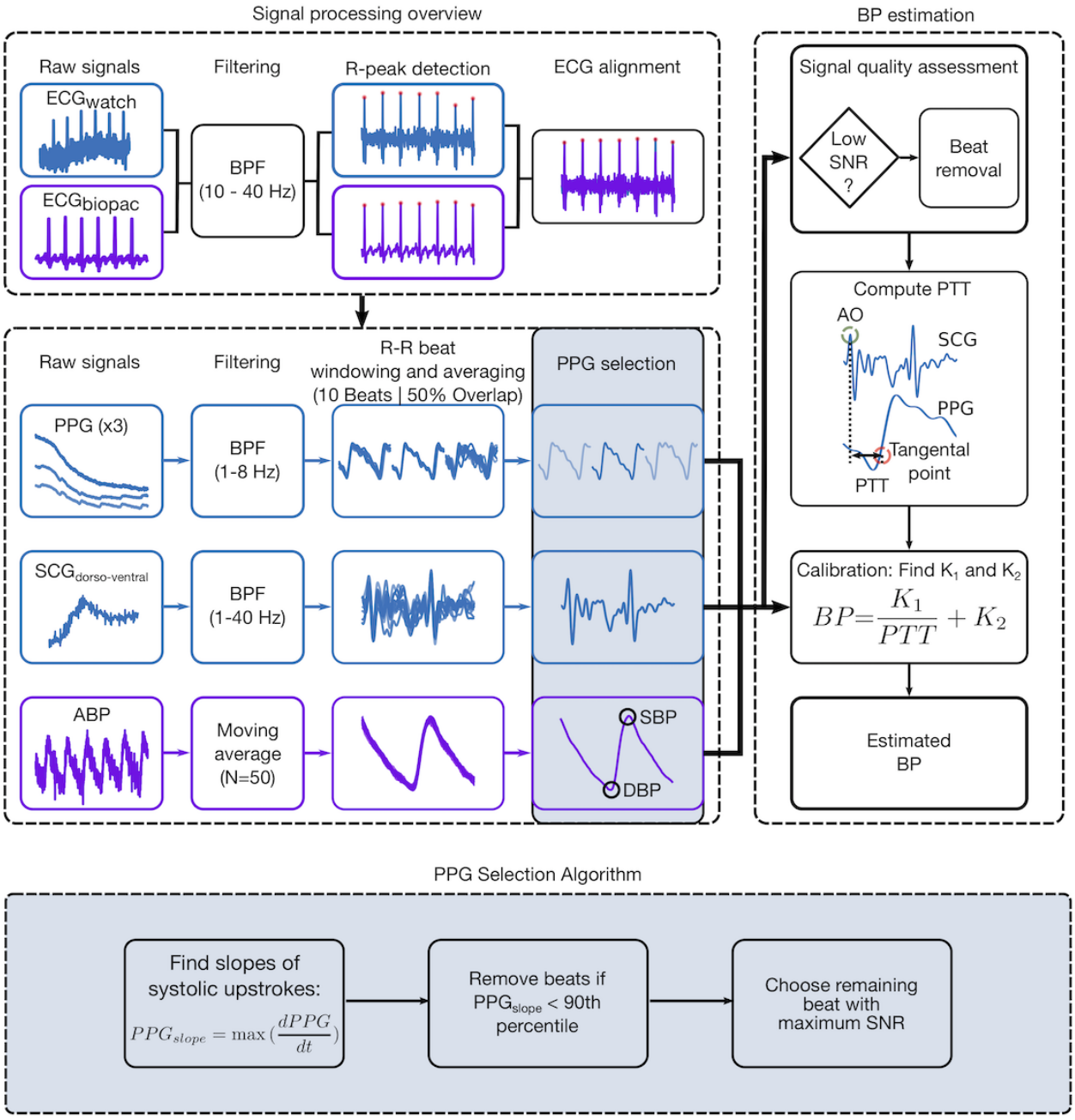
Participants were asked to change into either a V-cut T-shirt or tank top, if not wearing one already, to acquire the sternal PPGs included in the wearable designs, though not examined in this work. The watch was fitted such that the PPGs faced the radial artery on the ventral side of the wrist. To capture the PTT, the participant performed a simple maneuver to place the watch on the sternum to acquire the SCG for the proximal timing reference, as shown in Figure 1, whereas the PPGs were sampled at both the sternum and wrist. Although this offers a noncontinuous assessment, routine remote BP monitoring using oscillometric devices has already demonstrated clinical value, although it does not offer continuous BP measurement [2]. Specifically, ambulatory BP monitors, due to their superior portability and measurement frequency—comparable with what this wearable device can easily provide [20]—have become invaluable for the screening and management of hypertension [2] such that the added benefit of continuous BP measurement may only be marginal.

In order, the participants went through a 2-minute baseline period while sitting before obtaining another 2-minute baseline measurement while standing. Then, a series of perturbations with varying rest periods in between were used to modulate BP. First, a mental arithmetic exercise was used to increase BP [12], in which participants were given a three-digit integer and were told to add the sum of the digits to the number repeatedly for 1 minute. Then, a cold pressor test was conducted in which participants submerged their hand contralateral to the watch in a bucket of ice water for as long as tolerable or until the full minute. Finally, during the exercise session, the finger cuff was removed to avoid damage, and the participant performed either a stair stepping or bicycling exercise, based on personal preference, for 1 minute. As mentioned in our previous work [20], the new version of the watch enters the PTT measurement mode when the user places a finger from their hand contralateral to the watch on the positive wrist ECG electrode; therefore, we were unable to acquire PTT data during exercise for both cohorts and cold pressor for the participants in the second cohort (ie, second cohort). Although with the newer hardware, we were unable to collect PTT data during the cold pressor perturbation for the second cohort, the effect of the cold pressor—assessed directly after the hand was removed from the ice water (ie, a maximum of 1 minute after immersion)—was still well within its physiological window during the following rest period [25]. Overall, as our device is not designed to offer continuous measurements of BP, examining the effect of these perturbations in the rest period directly following them, similar to our previous work [19], still allowed for a comprehensive evaluation of the methodology in a diverse population. However, PTT data from the first cohort during the cold pressor were still used. As the BP data from the cold pressor test were still acquired for the second cohort, as the continuous BP cuff was still on, the mean arterial pressure (MAP) values were factored into the ones displayed in Figure 1. To do so, a 50 ms moving average filter was applied to the measured continuous BP signal, ensemble averages of 10 heartbeats with 50% overlap were taken, and the BP beat with the highest signal-to-noise ratio (SNR) was selected.

Signal Processing

The signal processing pipeline is shown in Figure 4. All signal processing and statistical analyses were performed in MATLAB R2018a (MathWorks). Before preprocessing the SCG and PPG signals acquired from the wearable system, it was imperative to time-align them to the continuous BP signal from the reference system using the respective ECGs to ensure proper temporal comparison. Specifically, the ECGs from each system were first filtered using a digital finite impulse response bandpass filter (BPF; fpass=10-40 Hz) to remove baseline wander due to postural sway and extract the R-wave, which was then identified using a simple peak detection algorithm. Then, cross-correlation was used to align the R-peaks of the two ECG readings by detecting the amount of lead and truncating either the wearable or reference signals depending on the condition. After alignment, the dorsoventral axis of the SCG (ie, z-axis acceleration) and green, red, and IR wrist PPGs were filtered using a digital finite impulse response BPF with bandwidths of 1-40 Hz and 1-8 Hz, respectively, to remove their out-of-band noise and baseline wander due to respiration. In addition, the continuous ABP waveform was smoothed using a 50 ms moving average filter.

Figure 4. Signal processing pipeline. Block diagram of signal processing overview showing signal alignment using electrocardiogram signals acquired from the wearable system (blue) and reference system (purple) before bandpass filtering, heartbeat windowing, and photoplethysmogram (PPG) selection. After beat selection and signal quality assessment, the pulse transit time is computed as the aortic valve opening point of the seismocardiogram to the diastolic foot of the PPG. Calibration is used to estimate blood pressure (BP) using the arterial blood pressure waveform acquired from the continuous BP finger-cuff. Block diagram of the custom PPG selection algorithm, locating beats with greater systolic upstrokes and signal-to-noise ratio (SNR). ABP: arterial blood pressure; AO: aortic valve opening; BP: blood pressure; BPF: bandpass filter; ECG: electrocardiogram; PPG: photoplethysmogram; PTT: pulse transit time; SBP: systolic blood pressure; SCG: seismocardiogram; SNR: signal-to-noise ratio.



Next, the filtered and aligned SCG, PPG, and ABP waveforms were split into separate heartbeats using the detected R-R intervals of the synchronized ECG. Then, these heartbeat-indexed signals were ensemble-averaged using 10-beat windows with 50% overlap before assessing the signal quality to select the highest quality beat per task for each participant, similar to the methods used in our previous works [19,20]. Given the number of high BMI participants in this population, the SCG not only had a lower mean SNR when compared with our previous studies but was also observed to have less variability than the PPG SNR; hence, an emphasis was placed on determining the optimal PPGs first. In addition, upon an initial assessment of signal quality, it was observed that when the PPG signal had the highest SNR, typically, the SCG signal did as well—perhaps because acquiring a clean PPG signal inherently hinges on applying consistent pressure. The optimum PPG was selected using a physiologically inspired algorithm to first identify the beats with the top 10% of systolic upstrokes (ie, maximum of the derivative of the PPG waveform) and then select the remaining beat with the maximum SNR. The SNR

was calculated using a noise-to-signal ratio detection algorithm detailed in Inan et al [26]. The methods used to determine the timing references for PTT calculation, the foot of the PPG, and the aortic valve opening (AO) point of the SCG were the same as those used in our previous studies [19,20]. Specifically, the foot of the PPG was computed using the intersecting tangent method described in the study by Mukkamala et al [12], and the AO point was assumed to be the first peak in each ensemble-averaged window before the foot of the PPG. Occasionally, the SCG signal was manually annotated to impose realistic constraints for the AO point or to ensure that the same morphological peak was consistently chosen for all tasks per participant. Participant-specific SNR thresholds were set to retain only high-fidelity signals; if the SNR of the SCG, PPG, or ABP beats was not greater than the prescribed cutoff, or if the foot of the PPG was not within a realistic range, then the respective ensemble-averaged waveforms were deemed too noisy for use and that task was not used for PTT calculation. Notably, the continuous reference BP allowed for the ability to evaluate the SNR of the ABP signal and incorporate this quality assessment into our signal processing pipeline to remove beats with low SNR reference measurements. After the entire signal quality assessment process, at least four of the tasks were used for BP estimation per participant. Finally, the PTT was calculated as the difference of the proximal timing reference, AO point of the paired SCG, and distal timing reference, the foot of the selected PPG. In addition to wavelength comparisons, the green and red wavelength PPGs were not used as the IR wavelength wrist PPGs had the highest mean SNR, because of the greater indifference of the IR wavelength to melanin absorption and the ability to capture more pulsatile arteries deeper in the tissue than cutaneous capillaries [12,27].

In addition, the postexercise recovery period was separated into an early and late rest period based on when the BP reached a consistent value. This allowed us to capture both the immediately heightened cardiac output–induced BP increase postexercise and the recovery back to baseline, while opportunely adding another PTT and BP data point for linear regression.

Statistical Analysis

Simple linear regression was performed independently between wearable participant-specific inverse PTT (PTT^{-1}) and reference diastolic blood pressure (DBP), MAP, or systolic blood pressure (SBP) value pairs, to calculate the calibration coefficients necessary to estimate each of the three BP components per participant; nonlinear models, whereas potentially more accurate, dictate the need for more calibration points [12,28]. Therefore, the resulting calibration coefficients—used to estimate BP from the conventional PTT-based BP estimation model shown in equation 1—are merely the slope (ie, slope calibration coefficient [K_1]) and y-intercept (ie, Y-intercept calibration coefficient [K_2]) of the line of best fit [12]. This was identical to the calibration methods used in our previous work [19,20].

$$BP = (K_1 / PTT) + K_2 \quad \textbf{(1)}$$

The mean absolute difference (MAD) was computed from the mean of the absolute value of the difference between the estimated and reference BP. The benchmarks for MAD were chosen based on the Institute for Electronics and Electrical Engineers (IEEE) standard for wearable cuffless BP measuring devices [29]. In addition, the root mean square error (RMSE), calculated as the root mean square of the difference between the estimated BP and measured BP, was computed because of its enhanced sensitivity to outliers.

We stratified the entire study population for the demographic comparisons of the calibration coefficients shown in Figure 3, based on four factors (ie, obesity, sex, race, and age) known to affect arterial stiffness [30-35] and therefore the PTT. The participants were split into nonobese and obese groups based on the guidelines from the National Heart, Lung, and Blood Institute of the National Institutes of Health defining a BMI ≥30 kg/m^2 as obese [21]. Thus, the nonobese group had a BMI ≤30. To assess differences due to age, we separated the participants into younger (aged ≤40 years) and older groups (aged ≥40 years). Statistical significance (P<.05) between demographic data for each cohort was assessed using an unpaired two-sample, two-tailed t test, as shown in Table 1.

For the demographic DBP calibration coefficient comparisons, a one-sample Kolmogorov-Smirnov test was used on each data point to test for normality, which determined that none of the data for the comparisons were normally distributed. Then, a Mann-Whitney U test (ie, Wilcoxon Rank Sum test) was used to assess statistical significance (P<.05) among the unpaired data.

For the PPG wavelength DBP estimation comparisons—only applicable to the second cohort population due to the differences in hardware used—first, a one-sample Kolmogorov-Smirnov test was used on each data point to test for normality, which determined that none of the data for the comparisons were normally distributed. Then, a Wilcoxon Signed Rank test was used to assess statistical significance (P<.05) among the paired data.

Results

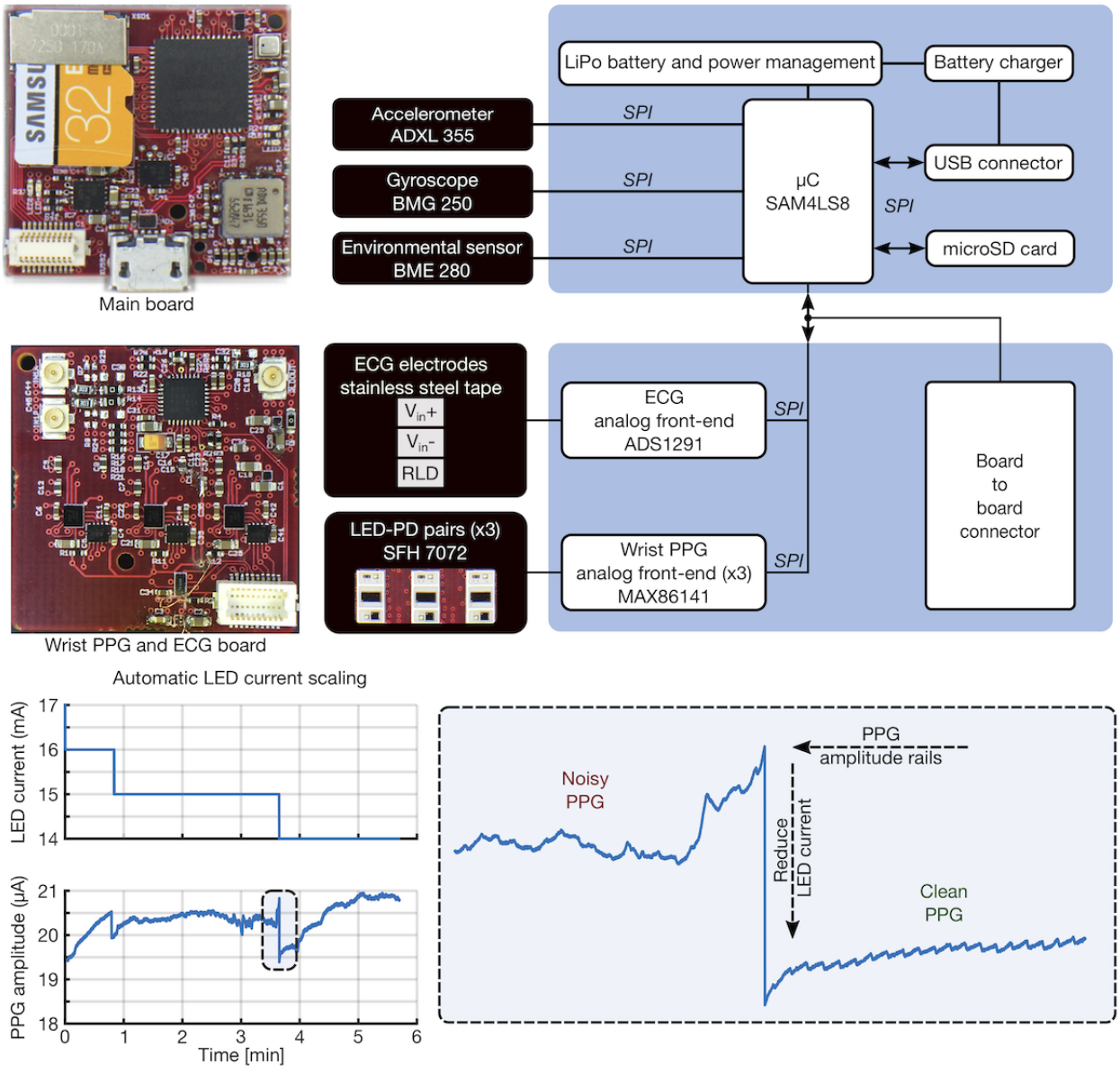
Multimodal Engineering Mechanics of the SeismoWatch

The previous version of the watch, not shown in this work, consisted of a 3D printed case embedded with an accelerometer, PDs, and IR LEDs. All sensors were connected to a small external circuit box with straps for the participant to wear around the waist. The output of the analog accelerometer (ADXL354, Analog Devices) was connected to an analog front end (AFE) in the circuit box. To amplify the SCG signal and prevent saturation of the alternating current components owing to the varying direct current levels, the AFE separated the direct current and alternating current components using a low pass (fc=1 Hz; G=−10 dB) and BPF (fpass=0.2 Hz-40 Hz) in parallel. An analog adder recombined both components. For PPG measurements, the cathode of the PDs (S2386-18k, Hamamatsu Photonics) was connected to transimpedance amplifiers configured as a low-pass filter (fc=12 Hz; G=110 dB) followed

by gain and filter stages (fpass=0.5-12 Hz; G=59 dB). Finally, the ECG was acquired by placing 3 copper dry electrodes on the wrist band of the watch with 2 on the inside in contact with the wrist and 1 on the outside to place the index and middle finger. The 2 on the inside act as the right leg drive electrode and the positive lead, whereas the outside electrode is the negative lead. All electrodes were connected to an AFE (AD8232, Analog Devices) for ECG measurements. A microcontroller (Teensy 3.6, PJRC LLC) sampled the output of the accelerometer, PPG, and ECG AFE at 1 kHz. An onboard SD card was used to store the raw data for postprocessing, and a 1.2 Ah lithium-ion rechargeable battery was used to power the system. All instrumentation details were adopted from our previous work, with minor revisions [19].

The updated hardware, pictured in Figure 5, added modalities of sensing (ie, a gyroscope), included multiple wavelengths of LEDs for comparison with IR, improved the form factor for comfort and ease of use, and featured embedded systems innovations leveraged in this study. A more thorough description of the revised hardware is available in our most recent work [20]. An example of the serviceable automatic LED current scaling algorithm, detailed in our previous work [20], is highlighted in Figure 5. This proved to be an integral part of enabling this work; by adaptively adjusting the LED drive current, we were able to prevent saturation and variable PPG signal quality caused by varying contact pressure and, more importantly, prominent differences in skin tone among participants.

Figure 5. Pertinent multimodal hardware block diagram and adaptive light-emitting diode (LED) scaling. Main board with ATSAM4LS8 microcontroller (µC), ADXL355 triaxial accelerometer, BMG250 triaxial gyroscope, and BME280 environmental sensor using the serial peripheral interface for fast communication supporting higher sample rates. Sensor board used to acquire wrist photoplethysmogram (PPG) and electrocardiogram signals. Automatic LED current scaling in operation during data collection: showing an increase in contact pressure and subsequent saturation of the photodiode, mitigated by an automatic decrease in LED current and overall consequential improvement in PPG signal quality. ECG: electrocardiogram; LED: light-emitting diode; PD: photodiode; PPG: photoplethysmogram; RLD: right leg drive; SD: Secure Digital; SPI: Serial Peripheral Interface.



Human Subject Studies in a Diverse Population

All applicable results are presented as mean (SD). Figure 2 illustrates the correlation and Bland-Altman plots for our wearable PTT-based BP estimation of MAP, DBP, and SBP across all participants (N=44). The MAD was 2.90 mm Hg, 3.39 mm Hg, and 5.36 mm Hg for DBP, MAP, and SBP, respectively. The mean RMSE was 3.41 (SD 2.01) mm Hg, 3.95 (SD 2.42) mm Hg, and 6.28 (SD 3.44) mm Hg for DBP, MAP, and SBP, respectively. DBP and MAP estimation had better 95% CIs than SBP at 7.99 mm Hg, 9.42 mm Hg, and 14.59 mm Hg, respectively. The mean Pearson correlation coefficient (PCC) was 0.67 (SD 0.16), 0.63 (SD 0.31), and 0.50 (SD 0.41) for PTT-based DBP, MAP, and SBP estimation, respectively.

The MAD for the individual study populations (first cohort=26 participants and second cohort=18 participants) was 2.69 mm Hg and 3.20 mm Hg, 3.21 mm Hg and 3.64 mm Hg, and 5.17 mm Hg and 5.63 mm Hg, for DBP, MAP, and SBP estimation, respectively. The mean RMSE for the individual study populations (first cohort=26 participants and second cohort=18 participants) was 3.19 (SD 1.64) mm Hg and 3.73 (SD 2.48) mm Hg, 3.78 (SD 2.06) mm Hg and 4.18 (SD 2.90) mmHg, and 6.26 (SD 3.25) mm Hg and 6.32 (SD 3.80) mm Hg for DBP, MAP, and SBP estimation, respectively. The mean PCC for the individual study populations (first cohort=26 participants and second cohort=18 participants) was 0.69 (SD 0.15) and 0.65 (SD 0.17), 0.68 (SD 0.23) and 0.55 (SD 0.38), and 0.58 (SD 0.33) and 0.39 (SD 0.49) for DBP, MAP, and SBP estimation, respectively.

The MAD for the 19 Black participants was 3.18 mm Hg, 3.72 mm Hg, and 5.84 mm Hg for DBP, MAP, and SBP estimation, respectively. The mean RMSE for all 19 Black participants was 3.72 (SD 2.41) mm Hg, 4.29 (SD 2.86) mm Hg, and 6.69 (SD 4.03) mm Hg for DBP, MAP, and SBP estimation, respectively. The mean PCC for all 19 Black participants was 0.64 (SD 0.17), 0.53 (SD 0.38), and 0.37 (SD 0.48) for DBP, MAP, and SBP estimation, respectively.

The MAD for the 10 participants who were obese was 2.69 mmHg, 3.17 mm Hg, and 5.02 mm Hg for DBP, MAP, and SBP estimation, respectively. The mean RMSE for all 10 participants who were obese was 3.28 (SD 2.59) mm Hg, 3.69 (SD 3.00) mm Hg, and 5.71 (SD 4.18) mm Hg for DBP, MAP, and SBP estimation, respectively. The mean PCC for all 10 participants who were obese was 0.65 (SD 0.18), 0.52 (SD 0.48), and 0.39 (SD 0.58) for DBP, MAP, and SBP estimation, respectively.

Figure 3 depicts the boxplots of the DBP calibration coefficients from our estimation model, K_1 and K_2, for four different demographic factors known to affect arterial stiffness: obesity, sex, race, and age. The DBP K_1 and K_2 values for nonobese (N=34) versus obese (N=10) participants are 2.38 (SD 1.99) mm Hg/s versus 1.20 (SD 0.88) mm Hg/s and 61.02 (SD 18.03) mm Hg versus 74.31 (SD 5.14) mm Hg, respectively. The DBP K_1 and K_2 values for male (N=25) versus female (N=19) participants are 2.65 (SD 2.18) mm Hg/s versus 1.40 (SD 0.98) mm Hg/s and 60.16 (SD 20.64) mm Hg versus 69.14 (SD 8.29) mm Hg, respectively. The DBP K_1 and K_2 values for non-Black (N=25) versus Black (N=19) participants are 2.63 (SD 2.21) mm Hg/s versus 1.44 (SD 0.94) mm Hg/s and 60.66 (SD 20.29) mm Hg versus 68.49 (SD 9.98) mm Hg, respectively. The DBP K_1 and K_2 values for young (N=31) versus older (N=13) participants are 2.38 (SD 2.09) mm Hg/s versus 1.47 (SD 0.87) mm Hg/s and 61.96 (SD 19.03) mm Hg versus 69.00 (SD 9.22) mm Hg, respectively.

Both K_1 and K_2 were significantly different between the nonobese and obese populations (P=.045 and P=.008, respectively). The female K_1 values were significantly (P=.04) lower than those of their male counterparts. The K_1 values for Black participants were significantly (P=.047) lower than those of the other races.

For the participants in the second cohort—all Black—with whom we used the newer version of the hardware [20] that included green and red LEDs in addition to the IR, the PCC for DBP estimation was 0.38 (SD 0.34), 0.59 (SD 0.44), and 0.65 (SD 0.17) when using the green λ=526 nm, red λ=660 nm, and IR λ=950 nm wavelength PPGs for the distal timing reference, respectively. The PCC for the IR and red wavelength PPGs was significantly (P=.01 and P=.048) higher than that of the green wavelength PPGs. However, the corresponding mean DBP RMSE was 3.95 (SD 2.53) mm Hg, 3.11 (SD 2.33) mm Hg, and 3.73 (SD 2.48) mm Hg for green, red, and IR, respectively.

Discussion

Principal Findings

To the best of our knowledge, this is the first study to accurately estimate DBP and MAP using noninvasive PTT measurements acquired from a holistic population, with considerable differences in body fat percentage, melanin levels, and vascular stiffness associated with age and hypertension. Furthermore, our SBP estimation is sufficient to be clinically recommended for monitoring [29,36]. We demonstrated the reliability of a convenient method for estimating BP and observed that our calibration coefficients were significantly different in characteristic demographic groups known to have increased arterial stiffness. This work represents a necessary advancement toward remote monitoring for persons in MUAs by enabling wearable PTT-based BP estimation, including through the comprehensive evaluation of a watch-based form factor conducive to obtaining ambulatory BP measurements in low-resource settings.

Accurately Estimating BP in a Diverse Population Using a Multimodal Wearable Device

We demonstrated the performance of our wrist-worn PTT-based device when used to estimate BP within a diverse population over the course of multiple unique perturbations. Our results for MAP and DBP passed the acceptable benchmarks for the BP estimation error set by the IEEE standard on wearable cuffless BP estimation devices (MAD≤5 mm Hg) [29]. We were still able to achieve a reliable correlation between PTT and BP even with several demographic factors such as age, melanin levels, and BMI inherently influencing the measured optical-PPG and mechanical-SCG signals.

The DBP estimation remained the most accurate, similar to our previous studies [19,20]; the foot of the PPG waveform, used as the distal timing reference, indicates the arrival of the pulse wave during end *diastole*. Similarly, the SBP estimation continued to perform the worst, as the peak of the pulse wave is the fiducial marker of the PPG that occurs during *systole*; however, the peak is not frequently extracted, as its true timing can be confounded by wave reflection interference, leading to unreliable PTT estimates [12]. Recent studies have demonstrated that the PTT computed using the diastolic foot of the PPG outperforms that using the systolic maximum for both DBP and SBP [37].

The DBP RMSE was relatively similar at both low and high values of DBP, which indicates that the diastolic foot was a dependable timing reference for calculating the PTT, irrespective of inherent participant-specific differences in BP. Although our SBP estimation was just outside the acceptable limits set forth by the IEEE standard (ie, MAD=5.36 mm Hg vs 5.0 mm Hg) [29], this error translates to a grade B classification [29] and therefore would still be clinically recommended for monitoring SBP [36]. Furthermore, the SBP range studied in this work was greater than 100 mm Hg, substantially higher than that reported in previous studies in the literature for wearable cuffless BP estimation, and a combination of different perturbations was used to modulate BP. Using a single perturbation would have led to an improved correlation [12,14], as in our previous work where we had only used exercise [19]. However, a comprehensive evaluation of this methodology would be incomplete without a procedure consisting of a wide variety of perturbations with different known physiological responses and pathways to modulate BP [14]. In addition, as noted in Figure 1, the exercise perturbation did not apparently produce a marked difference in BP due to several factors: (1) technical limitations in rapid calibration for the reference measurement (ie, finger-cuff continuous BP) and increased motion artifacts following exercise led to a greater percentage of beat removal in the early exercise section than any other task and (2) exercise does not necessarily consistently modulate BP in a predictable manner due to differences in participant-specific vasoactivity and contractility [12,38].

Only the DBP was examined for further analyses conducted below because, as previously mentioned, the distal timing reference used in this work (ie, the foot of the PPG waveform) occurs during diastole and therefore provides the most reliable estimation of DBP out of the three BP components [12]. The dependability of the diastolic foot and our robust DBP estimation were necessary before performing in-depth analyses with confidence. Although elevated SBP is considered to be the greatest predictor of future cardiovascular risk [39,40], elevated DBP has nonetheless been shown to independently increase the risk of subsequent cardiac events [39,41]. In addition, DBP is a greater contributor to MAP, which in older patients with isolated systolic hypertension, when compared with an equivalent increase in pulse pressure, has been shown to be a comparable independent predictor of both stroke and all-cause mortality [42]. Finally, DBP has been shown to be a more significant predictor than SBP of new-onset hypertension in adults aged ≤50 years [40,43-45]. This suggests that accurate DBP estimation using a wearable device can efficiently be used to incentivize people to make healthy lifestyle modifications *earlier in life*, central to the World Health Organization's effort to reduce the global prevalence of hypertension [46].

Essential Device Novelties Enabling Reliable PTT Computation

For the first time, we demonstrated that noninvasive PTT measurements are reliable estimators of BP across a wide range of skin tones and BMI. Both DBP and MAP estimation for the 10 participants who are obese and 19 Black participants in this study were well under the IEEE requirement [29]. This was enabled by the highly sensitive hardware, multisensor approach, and automated LED current scaling that our custom wearable device offers [20]. The PPG array and adaptive LED current scaling allowed us to automatically mitigate poor signal quality issues due to misplacement, inherent differences in skin tone, and applied pressure that typically corrupts PPG signals. However, the most integral components of our PPG hardware were the IR wavelength LEDs.

We leveraged longer wavelengths of light for deeper penetration into the tissue to robustly acquire the PPG signal from arteries located deeper than the cutaneous vascular bed [12]. Cutaneous arteries are greatly affected by the changes in vascular tone expected from the perturbations we used to modulate BP herein (ie, cold pressor and exercise). Furthermore, as IR PPGs are more susceptible to motion artifacts than lower wavelength ones [12,47], our PPG-first signal quality assessment not only avoided these motion artifact corrupted waveforms because of their low SNR but also avoided moments where the SCG quality would naturally suffer as well. However, even the red PPGs had a considerably larger SD in their PCC than the IR PPGs, possibly because the IR wavelength, when compared with red, is less sensitive to the oxygen content of hemoglobin and has approximately half the skin absorption coefficient in Black individuals [12,27]. Despite statistically significant differences in the PCC using IR and red PPGs rather than green PPGs, the actual DBP RMSEs were comparable. This implies that when using the green PPGs for participants with a low PCC, our signal quality assessment algorithm removed beats with greater BP variability, resulting in a lower SD of DBP and consequent RMSE. Although even green wavelength PPGs have demonstrated the ability to reliably extract heart rate across a wide variety of skin tones [48], our data suggest that these shorter wavelengths cannot be used to dependably compute the PTT in a diverse population. In addition, although unconventional, our watch was placed on the ventral side of the wrist, which allows for both higher quality, convenient SCG acquisition and enhanced PPG SNR due to viable access to the radial artery and less melanin content than the dorsal side [49]. Therefore, existing smartwatches, beginning to slowly incorporate cuffless, noninvasive BP methodologies, may face even greater difficulties in achieving accurate PPG measurements across a broad range of skin tones.

Finally, our physiologically inspired PPG selection algorithm—to first select the PPG signals with the greatest systolic upstrokes—had an important role in reducing the BP estimation error. PPG waveforms with greater systolic upstrokes

(ie, maximum derivative of the PPG waveform) offer improved PTT estimates and are key indicators of BP stemming from larger, more pulsatile, elastic arteries with greater distensibility [12,50]. In addition, several recent machine learning (ML) approaches to use the PPG signal for BP estimation have shown that the systolic upstroke is one of the most important features of the waveform [51,52]. Hence, the selection algorithm, by extracting information from these more reliable and clinically important arteries, was a central part of our ability to notice the demographic differences in arterial stiffness rooted in our calibration coefficients.

Calibration Coefficients Capture Demographic Differences in Arterial Stiffness

We observed that the participant-specific calibration coefficients used in the standard linear PTT-BP estimation model for DBP, shown in equation 1 (ie, K_1 and K_2), are significantly different between subpopulations with large variations in demographic factors known to affect arterial stiffness. We selected the four demographic categories (ie, obesity, sex, race, and age) based on the literature, emphasizing these as major determinants of differences in arterial stiffness and therefore risk factors for hypertension [16,30-33,35,53,54].

The K_1 value (ie, the slope of the line of best fit) is indicative of the underlying baseline vascular stiffness, whereas K_2 (ie, the intercept) represents the inherently correlated bias in baseline BP [12,55,56]. At the same BP, persons with greater arterial stiffness have inherently faster pulse wave velocities (PWVs) and therefore shorter PTTs than persons with normal arterial stiffness [12]. The K_1 value mitigates these differences in PTT-based estimation by capturing the intrinsic participant-specific arterial stiffness to output similar B*P* values. Therefore, with increasing arterial stiffness, we expected to find a lower K_1 value and a higher K_2 value, as observed in the PWV literature [55,56].

Obesity was the only comparison for which the differences in the K_1 and K_2 calibration coefficients were statistically significant. This coincides with the literature stating that obesity is one of the greatest age-normalized risk factors and contributors to arterial stiffness [57]. Otherwise, only the K_1 values in the sex and race comparisons were statistically significant between the groups. Although it has been shown that both females and Black individuals have greater arterial stiffness than similar-age males and White individuals [31,34,35], these two comparisons should be re-evaluated after increasing our recruitment. Approximately 47% (9/19 participants) of both the female and Black population were also obese. The age comparison was not statistically significant, although the older population followed a similar trend of a lower K_1 and higher K_2. This finding is not surprising, as significant differences in arterial stiffness and substantial augmentations in arterial remodeling are typically examined in participants aged ≥50 years [32,58].

Limitations and Future Work

Refining Population Demographics and Investigating PTT, K1, and K2 as Potential Digital Biomarkers of Arterial Stiffness

Overall, although this data set captured a more representative population in the range of end users for which consistent BP monitoring is recommended [59], our PTT-based device should be further tested in an exclusively older (ie, age >50 years), morbidly obese (ie, BMI >40 kg/m^2), and hypertensive population—with even distributions across sex, race, and skin tones along the Fitzpatrick scale—to truly understand the limits of this technology and supplement the findings herein.

Early vascular remodeling due to the demographic factors investigated in this work, not to mention socioeconomic factors affecting MUAs [15,16], predispose individuals who are obese and Black individuals to greater lifetime cardiovascular risk [30,35,57,60,61]. Therefore, future PTT-based BP estimation studies should closely monitor the calibration coefficients, K_1 and K_2, as potential intermediate digital biomarkers for longitudinal monitoring and the comparison of arterial stiffness among different persons [7]. Eventually, even PTT measurements, as PWV is already an independent predictor of arterial stiffness [62], may indicate subclinical differences in vascular resistance due to early stage arterial remodeling, the main precursor to hypertension [32].

Reducing the Burden of Calibration

Consistent recalibration poses a practical concern for PTT-based BP estimation. Hence, future studies should focus on evaluating the timeframe for which participant-specific calibration curves can reliably estimate BP and whether interparticipant and population-level curves can be sufficient. However, given the value of interpreting the calibration coefficients presented in this work, caution should be exercised due to the trade-off of sacrificing this potential usefulness when using generalized interparticipant models. Furthermore, the individual effects of the perturbations used to modulate BP in this experiment should be scrutinized, along with other exercises shown to substantially change BP [63-65]. The goal is to use perturbations that can consistently be leveraged to increase the dynamic range of BP measurements for calibration—critical to achieving optimal estimations at home in our previous work [20] and are achievable in low-resource settings.

Leveraging ML and Hardware Advancements for Robust SCG AO Detection

Similarly, to the instrumental role of the physiologically inspired PPG selection algorithm in this work, further exploration into automated SCG fiducial point detection algorithms may help extract the most informative SCG signals. Specifically, the SCG can be greatly affected by inaccurate placement of the watch; however, recent advancements using ML techniques have shown that the SCG waveform is modulated in a predictable manner during these placement inaccuracies [66]. Therefore, by interpreting these findings, one might be able to convert the measured SCG to the archetypal SCG or use a template-matching localization approach [67] for each

participant before extracting salient features from the optimal waveform.

In addition, annotating the exact AO point can be challenging because the signal not only has appreciable interparticipant variability, especially in a population with considerable differences in BMI, but can also be corrupted by motion artifacts. Although our technique for extracting the AO point has led to a high correlation between PTT and BP, in both our recent work [20] and this one, for a few sessions, we manually annotated the SCG to impose realistic constraints for the range of the pre-ejection period (PEP) and selected a consistent morphological peak across all tasks per participant. Eventually, robust identification of this timing reference is necessary for reliable automatic PTT computation, as the main advantage of using the PTT over the pulse arrival time (ie, the time from the R-wave of the ECG to the diastolic foot of the PPG) for BP estimation is its ability to account for changes in the nonnegligible cardio-electromechanical delay, that is, the PEP [12,68]. Furthermore, examining the other sensor data available at our disposal, such as filtering the SCG in a higher bandwidth (ie, fpass=30-125 Hz) to retain the phonocardiogram signal indicative of valve closures, using the three-axis gyrocardiogram or simply the other axes of the SCG, could prove to help with improving PEP estimation as shown in previous work [19,69].

Conclusions

We have demonstrated that a wrist-worn device, using noninvasive PTT estimates, can reliably and conveniently track BP in a diverse population. Leveraging the ubiquity of wearable devices can empower users to make healthy lifestyle modifications such as exercise, which can contribute to a significant reduction in arterial stiffness [30,70] by providing consistent feedback on progress [71-73]. In addition, digital health technologies that accurately estimate BP could potentially be used to titrate BP medications for patients with hypertension from the comfort of their homes [7,74]. In addition to these broader impacts, the knowledge gained from this study—especially when combined with the advent of low-profile, flexible electronics capable of robustly detecting physiological biosignals [75-78]—represents a significant step toward the unobtrusive monitoring of BP in ambulatory settings and health equity for persons in MUAs.

Acknowledgments

This work was funded in part by the National Institute of Health, National Institute of Biomedical Imaging and Bioengineering under grant 1U01EB018818-01, by the National Institute of Biomedical Imaging and Bioengineering Point-of-Care Technologies Research Network, Atlanta Center for Microsystems-Engineered Point-of-Care Technologies under grant 1U54EB027690, and by the National Center for Advancing Translational Sciences of the National Institutes of Health under award number UL1TR002378.

Authors' Contributions

VGG codeveloped the newer version of the hardware, performed human subject studies on the second cohort, analyzed the collected data, and coprepared the manuscript. AMC developed the older version of the hardware, codeveloped the newer version of the hardware, co-conducted human subject studies on the first cohort, and assisted in advising the analysis and editing of the manuscript. HJ co-conducted human subject studies on the first cohort and assisted in both human subject studies on the second cohort and editing of the manuscript. AVS assisted in both conducting human subject studies on the second cohort and editing of the manuscript. DC assisted in both participant recruitment for the second cohort and editing of the manuscript. LNJ assisted in advising the study, participant recruitment for the second cohort, and editing of the manuscript. OTI guided the study and coprepared the manuscript.

Conflicts of Interest

OTI is a cofounder of and scientific advisor at Cardiosense, Inc, and a scientific advisor at Physiowave, Inc. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Arrieta A, Woods JR, Qiao N, Jay SJ. Cost-benefit analysis of home blood pressure monitoring in hypertension diagnosis and treatment: an insurer perspective. Hypertension 2014 Oct;64(4):891-896. [doi: 10.1161/HYPERTENSIONAHA.114.03780] [Medline: 25024284]
2. Ghuman N, Campbell P, White WB. Role of ambulatory and home blood pressure recording in clinical practice. Curr Cardiol Rep 2009 Nov;11(6):414-421 [FREE Full text] [doi: 10.1007/s11886-009-0060-6] [Medline: 19863865]
3. Yarows SA, Staessen JA. How to use home blood pressure monitors in clinical practice. Am J Hypertens 2002 Jan;15(1 Pt 1):93-96. [doi: 10.1016/s0895-7061(01)02277-4] [Medline: 11824867]
4. Muntner P, Hardy ST, Fine LJ, Jaeger BC, Wozniak G, Levitan EB, et al. Trends in blood pressure control among US adults with hypertension, 1999-2000 to 2017-2018. J Am Med Assoc 2020 Sep 22;324(12):1190-1200. [doi: 10.1001/jama.2020.14545] [Medline: 32902588]

5. Zhou D, Xi B, Zhao M, Wang L, Veeranki SP. Uncontrolled hypertension increases risk of all-cause and cardiovascular disease mortality in US adults: the NHANES III Linked Mortality Study. Sci Rep 2018 Jun 20;8(1):9418 [FREE Full text] [doi: 10.1038/s41598-018-27377-2] [Medline: 29925884]
6. Doyle SK, Chang AM, Levy P, Rising KL. Achieving health equity in hypertension management through addressing the social determinants of health. Curr Hypertens Rep 2019 Jun 12;21(8):58. [doi: 10.1007/s11906-019-0962-7] [Medline: 31190099]
7. Inan OT, Tenaerts P, Prindiville SA, Reynolds HR, Dizon DS, Cooper-Arnold K, et al. Digitizing clinical trials. NPJ Digit Med 2020;3:101 [FREE Full text] [doi: 10.1038/s41746-020-0302-y] [Medline: 32821856]
8. Sharma M, Barbosa K, Ho V, Griggs D, Ghirmai T, Krishnan S, et al. Cuff-less and continuous blood pressure monitoring: a methodological review. Technologies 2017 May 09;5(2):21. [doi: 10.3390/technologies5020021]
9. Arakawa T. Recent research and developing trends of wearable sensors for detecting blood pressure. Sensors (Basel) 2018 Aug 23;18(9):2772 [FREE Full text] [doi: 10.3390/s18092772] [Medline: 30142931]
10. Kario K, Shimbo D, Tomitani N, Kanegae H, Schwartz JE, Williams B. The first study comparing a wearable watch-type blood pressure monitor with a conventional ambulatory blood pressure monitor on in-office and out-of-office settings. J Clin Hypertens (Greenwich) 2020 Feb;22(2):135-141 [FREE Full text] [doi: 10.1111/jch.13799] [Medline: 31955505]
11. Bard DM, Joseph JI, van Helmond N. Cuff-less methods for blood pressure telemonitoring. Front Cardiovasc Med 2019;6:40 [FREE Full text] [doi: 10.3389/fcvm.2019.00040] [Medline: 31157236]
12. Mukkamala R, Hahn J, Inan OT, Mestha LK, Kim C, Töreyin H, et al. Toward ubiquitous blood pressure monitoring via pulse transit time: theory and practice. IEEE Trans Biomed Eng 2015 Aug;62(8):1879-1901 [FREE Full text] [doi: 10.1109/TBME.2015.2441951] [Medline: 26057530]
13. Etemadi M, Inan OT. Wearable ballistocardiogram and seismocardiogram systems for health and performance. J Appl Physiol (1985) 2018 Feb 01;124(2):452-461 [FREE Full text] [doi: 10.1152/japplphysiol.00298.2017] [Medline: 28798198]
14. Huynh TH, Jafari R, Chung W. Noninvasive cuffless blood pressure estimation using pulse transit time and impedance plethysmography. IEEE Trans Biomed Eng 2019 Apr;66(4):967-976. [doi: 10.1109/TBME.2018.2865751] [Medline: 30130167]
15. Bove AA, Santamore WP, Homko C, Kashem A, Cross R, McConnell TR, et al. Reducing cardiovascular disease risk in medically underserved urban and rural communities. Am Heart J 2011 Feb;161(2):351-359. [doi: 10.1016/j.ahj.2010.11.008] [Medline: 21315219]
16. Lincoln KD, Abdou CM, Lloyd D. Race and socioeconomic differences in obesity and depression among Black and non-Hispanic White Americans. J Health Care Poor Underserved 2014 Feb;25(1):257-275 [FREE Full text] [doi: 10.1353/hpu.2014.0038] [Medline: 24509025]
17. Saeed A, Dixon DL, Yang E. Racial disparities in hypertension prevalence and management: a crisis control? American College of Cardiology. 2020. URL: https://www.acc.org/latest-in-cardiology/articles/2020/04/06/08/53/racial-disparities-in-hypertension-prevalence-and-management [accessed 2021-06-23]
18. Lackland DT. Racial differences in hypertension: implications for high blood pressure management. Am J Med Sci 2014 Aug;348(2):135-138 [FREE Full text] [doi: 10.1097/MAJ.0000000000000308] [Medline: 24983758]
19. Carek AM, Conant J, Joshi A, Kang H, Inan OT. SeismoWatch: wearable cuffless blood pressure monitoring using pulse transit time. Proc ACM Interact Mob Wearable Ubiquitous Technol 2017 Sep;1(3):40 [FREE Full text] [doi: 10.1145/3130905] [Medline: 30556049]
20. Ganti VG, Carek A, Nevius BN, Heller J, Etemadi M, Inan O. Wearable cuff-less blood pressure estimation at home via pulse transit time. IEEE J Biomed Health Inform 2020 Sep 03;25(6):1926-1937. [doi: 10.1109/JBHI.2020.3021532] [Medline: 32881697]
21. The Practical Guide : identification, evaluation, and treatment of overweight and obesity in adults. NHLBI Obesity Education Initiative. 2000. URL: https://www.nhlbi.nih.gov/files/docs/guidelines/prctgd_c.pdf [accessed 2021-06-23]
22. Parati G, Casadei R, Groppelli A, Di Rienzo M, Mancia G. Comparison of finger and intra-arterial blood pressure monitoring at rest and during laboratory testing. Hypertension 1989 Jun;13(6 Pt 1):647-655. [doi: 10.1161/01.hyp.13.6.647] [Medline: 2500393]
23. Wesseling KH, de Wit B, van der Hoeven GM, van Goudoever J, Settels JJ. Physiocal, calibrating finger vascular physiology for finapres. Homeostasis in Health and Disease 1995;36(2-3):67-82 [FREE Full text]
24. Chachula K, Lieb F, Hess F, Welter J, Graf N, Dullenkopf A. Non-invasive continuous blood pressure monitoring (ClearSight™ system) during shoulder surgery in the beach chair position: a prospective self-controlled study. BMC Anesthesiol 2020 Oct 24;20(1):271 [FREE Full text] [doi: 10.1186/s12871-020-01185-6] [Medline: 33099306]
25. Zhao Q, Bazzano LA, Cao J, Li J, Chen J, Huang J, et al. Reproducibility of blood pressure response to the cold pressor test: the GenSalt Study. Am J Epidemiol 2012 Oct 01;176 Suppl 7:91-98 [FREE Full text] [doi: 10.1093/aje/kws294] [Medline: 23035148]
26. Inan OT, Kovacs GT, Giovangrandi L. Evaluating the lower-body electromyogram signal acquired from the feet as a noise reference for standing ballistocardiogram measurements. IEEE Trans Inf Technol Biomed 2010 Sep;14(5):1188-1196. [doi: 10.1109/TITB.2010.2044185] [Medline: 20371416]

27. Tseng S, Grant A, Durkin AJ. In vivo determination of skin near-infrared optical properties using diffuse optical spectroscopy. J Biomed Opt 2008;13(1):014016 [FREE Full text] [doi: 10.1117/1.2829772] [Medline: 18315374]
28. Yavarimanesh M, Chandrasekhar A, Hahn J, Mukkamala R. Commentary: Relation between blood pressure and pulse wave velocity for human arteries. Front Physiol 2019;10:1179 [FREE Full text] [doi: 10.3389/fphys.2019.01179] [Medline: 31572224]
29. 1708-2014 - IEEE Standard for Wearable Cuffless Blood Pressure Measuring Devices. Piscataway, New Jersey, United States: IEEE; 2014.
30. Wildman RP, Farhat GN, Patel AS, Mackey RH, Brockwell S, Thompson T, et al. Weight change is associated with change in arterial stiffness among healthy young adults. Hypertension 2005 Feb;45(2):187-192. [doi: 10.1161/01.HYP.0000152200.10578.5d] [Medline: 15596570]
31. Ahimastos AA, Formosa M, Dart AM, Kingwell BA. Gender differences in large artery stiffness pre- and post puberty. J Clin Endocrinol Metab 2003 Nov;88(11):5375-5380. [doi: 10.1210/jc.2003-030722] [Medline: 14602776]
32. Wu S, Jin C, Li S, Zheng X, Zhang X, Cui L, et al. Aging, arterial stiffness, and blood pressure association in Chinese adults. Hypertension 2019 Apr;73(4):893-899. [doi: 10.1161/HYPERTENSIONAHA.118.12396] [Medline: 30776974]
33. Alecu C, Gueguen R, Aubry C, Salvi P, Perret-Guillaume C, Ducrocq X, et al. Determinants of arterial stiffness in an apparently healthy population over 60 years. J Hum Hypertens 2006 Oct;20(10):749-756. [doi: 10.1038/sj.jhh.1002072] [Medline: 16855622]
34. Coutinho T. Arterial stiffness and its clinical implications in women. Can J Cardiol 2014 Jul;30(7):756-764. [doi: 10.1016/j.cjca.2014.03.020] [Medline: 24970788]
35. Schutte AE, Kruger R, Gafane-Matemane LF, Breet Y, Strauss-Kruger M, Cruickshank JK. Ethnicity and arterial stiffness. Arterioscler Thromb Vasc Biol 2020 May;40(5):1044-1054. [doi: 10.1161/ATVBAHA.120.313133] [Medline: 32237903]
36. Ding X, Zhang Y. Pulse transit time technique for cuffless unobtrusive blood pressure measurement: from theory to algorithm. Biomed Eng Lett 2019 Feb;9(1):37-52 [FREE Full text] [doi: 10.1007/s13534-019-00096-x] [Medline: 30956879]
37. Yang C, Tavassolian N. Pulse transit time measurement using seismocardiogram, photoplethysmogram, and acoustic recordings: evaluation and comparison. IEEE J Biomed Health Inform 2018 May;22(3):733-740. [doi: 10.1109/JBHI.2017.2696703] [Medline: 28436909]
38. MacDonald JR. Potential causes, mechanisms, and implications of post exercise hypotension. J Hum Hypertens 2002 Apr;16(4):225-236. [doi: 10.1038/sj.jhh.1001377] [Medline: 11967715]
39. Flint AC, Conell C, Ren X, Banki NM, Chan SL, Rao VA, et al. Effect of systolic and diastolic blood pressure on cardiovascular outcomes. N Engl J Med 2019 Jul 18;381(3):243-251. [doi: 10.1056/NEJMoa1803180] [Medline: 31314968]
40. Kannel WB, Gordon T, Schwartz MJ. Systolic versus diastolic blood pressure and risk of coronary heart disease. The Framingham study. Am J Cardiol 1971 Apr;27(4):335-346. [doi: 10.1016/0002-9149(71)90428-0] [Medline: 5572576]
41. Kaneko H, Itoh H, Yotsumoto H, Kiriyama H, Kamon T, Fujiu K, et al. Association of isolated diastolic hypertension based on the cutoff value in the 2017 American College of Cardiology/American Heart Association blood pressure guidelines with subsequent cardiovascular events in the general population. J Am Heart Assoc 2020 Oct 20;9(19):e017963 [FREE Full text] [doi: 10.1161/JAHA.120.017963] [Medline: 32993440]
42. Domanski MJ, Davis BR, Pfeffer MA, Kastantin M, Mitchell GF. Isolated systolic hypertension : prognostic information provided by pulse pressure. Hypertension 1999 Sep;34(3):375-380. [doi: 10.1161/01.hyp.34.3.375] [Medline: 10489379]
43. Kanegae H, Oikawa T, Okawara Y, Hoshide S, Kario K. Which blood pressure measurement, systolic or diastolic, better predicts future hypertension in normotensive young adults? J Clin Hypertens (Greenwich) 2017 Jun;19(6):603-610 [FREE Full text] [doi: 10.1111/jch.13015] [Medline: 28444926]
44. Franklin SS, Larson MG, Khan SA, Wong ND, Leip EP, Kannel WB, et al. Does the relation of blood pressure to coronary heart disease risk change with aging? The Framingham Heart Study. Circulation 2001 Mar 06;103(9):1245-1249. [doi: 10.1161/01.cir.103.9.1245] [Medline: 11238268]
45. Taylor BC, Wilt TJ, Welch HG. Impact of diastolic and systolic blood pressure on mortality: implications for the definition of "normal". J Gen Intern Med 2011 Jul;26(7):685-690 [FREE Full text] [doi: 10.1007/s11606-011-1660-6] [Medline: 21404131]
46. World Health Organization. Global Action Plan for the Prevention and Control of NCDs 2013-2020. Geneva: World Health Organization; 2013:1-103.
47. Convertino VA, Schauer SG, Weitzel EK, Cardin S, Stackle ME, Talley MJ, et al. Wearable sensors incorporating compensatory reserve measurement for advancing physiological monitoring in critically injured trauma patients. Sensors (Basel) 2020 Nov 10;20(22):6413 [FREE Full text] [doi: 10.3390/s20226413] [Medline: 33182638]
48. Bent B, Goldstein BA, Kibbe WA, Dunn JP. Investigating sources of inaccuracy in wearable optical heart rate sensors. NPJ Digit Med 2020;3:18 [FREE Full text] [doi: 10.1038/s41746-020-0226-6] [Medline: 32047863]
49. Tseng S, Bargo P, Durkin A, Kollias N. Chromophore concentrations, absorption and scattering properties of human skin in-vivo. Opt Express 2009 Aug 17;17(17):14599-14617 [FREE Full text] [doi: 10.1364/oe.17.014599] [Medline: 19687939]
50. Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiol Meas 2007 Mar;28(3):1-39. [doi: 10.1088/0967-3334/28/3/R01] [Medline: 17322588]

51. El-Hajj C, Kyriacou P. A review of machine learning techniques in photoplethysmography for the non-invasive cuff-less measurement of blood pressure. Biomed Signal Process Control 2020 Apr;58:101870. [doi: 10.1016/j.bspc.2020.101870]
52. Wang L, Zhou W, Xing Y, Zhou X. A novel neural network model for blood pressure estimation using photoplethesmography without electrocardiogram. J Healthc Eng 2018;2018:7804243 [FREE Full text] [doi: 10.1155/2018/7804243] [Medline: 29707186]
53. He J, Muntner P, Chen J, Roccella EJ, Streiffer RH, Whelton PK. Factors associated with hypertension control in the general population of the United States. Arch Intern Med 2002 May 13;162(9):1051-1058. [doi: 10.1001/archinte.162.9.1051] [Medline: 11996617]
54. Shea S, Misra D, Ehrlich MH, Field L, Francis CK. Predisposing factors for severe, uncontrolled hypertension in an inner-city minority population. N Engl J Med 1992 Sep 10;327(11):776-781. [doi: 10.1056/NEJM199209103271107] [Medline: 1501654]
55. Komine H, Asai Y, Yokoi T, Yoshizawa M. Non-invasive assessment of arterial stiffness using oscillometric blood pressure measurement. Biomed Eng Online 2012 Feb 10;11:6 [FREE Full text] [doi: 10.1186/1475-925X-11-6] [Medline: 22325084]
56. Hughes DJ, Babbs CF, Geddes LA, Bourland JD. Measurements of Young's modulus of elasticity of the canine aorta with ultrasound. Ultrason Imaging 1979 Oct;1(4):356-367. [doi: 10.1177/016173467900100406] [Medline: 575833]
57. Hall JE, do Carmo JM, da Silva AA, Wang Z, Hall ME. Obesity-induced hypertension: interaction of neurohumoral and renal mechanisms. Circ Res 2015 Mar 13;116(6):991-1006 [FREE Full text] [doi: 10.1161/CIRCRESAHA.116.305697] [Medline: 25767285]
58. AlGhatrif M, Strait JB, Morrell CH, Canepa M, Wright J, Elango P, et al. Longitudinal trajectories of arterial stiffness and the role of blood pressure: the Baltimore Longitudinal Study of Aging. Hypertension 2013 Nov;62(5):934-941 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.113.01445] [Medline: 24001897]
59. Sinha MD, Gilg JA, Kerecuk L, Reid CJ, British Association for Paediatric Nephrology. Progression to hypertension in non-hypertensive children following renal transplantation. Nephrol Dial Transplant 2012 Jul;27(7):2990-2996. [doi: 10.1093/ndt/gfr784] [Medline: 22287656]
60. Scuteri A, Orru' M, Morrell CH, Tarasov K, Schlessinger D, Uda M, et al. Associations of large artery structure and function with adiposity: effects of age, gender, and hypertension. The SardiNIA Study. Atherosclerosis 2012 Mar;221(1):189-197 [FREE Full text] [doi: 10.1016/j.atherosclerosis.2011.11.045] [Medline: 22222417]
61. Leggio M, Lombardi M, Caldarone E, Severi P, D'Emidio S, Armeni M, et al. The relationship between obesity and hypertension: an updated comprehensive overview on vicious twins. Hypertens Res 2017 Dec;40(12):947-963. [doi: 10.1038/hr.2017.75] [Medline: 28978986]
62. Najjar SS, Scuteri A, Shetty V, Wright JG, Muller DC, Fleg JL, et al. Pulse wave velocity is an independent predictor of the longitudinal increase in systolic blood pressure and of incident hypertension in the Baltimore Longitudinal Study of Aging. J Am Coll Cardiol 2008 Apr 08;51(14):1377-1383 [FREE Full text] [doi: 10.1016/j.jacc.2007.10.065] [Medline: 18387440]
63. Philips JC, Marchand M, Scheen AJ. Squatting, a posture test for studying cardiovascular autonomic neuropathy in diabetes. Diabetes Metab 2011 Dec;37(6):489-496. [doi: 10.1016/j.diabet.2011.09.004] [Medline: 22071282]
64. Rossberg F, Penaz J. Initial cardiovascular response on change of posture from squatting to standing. Eur J Appl Physiol Occup Physiol 1988;57(1):93-97. [doi: 10.1007/BF00691245] [Medline: 3342800]
65. Barnes SC, Ball N, Haunton VJ, Robinson TG, Panerai RB. The cerebrocardiovascular response to periodic squat-stand maneuvers in healthy subjects: a time-domain analysis. Am J Physiol Heart Circ Physiol 2017 Dec 01;313(6):1240-1248 [FREE Full text] [doi: 10.1152/ajpheart.00331.2017] [Medline: 28887332]
66. Zia J, Kimball J, Hersek S, Inan OT. Modeling consistent dynamics of cardiogenic vibrations in low-dimensional subspace. IEEE J Biomed Health Inform 2020 Jul;24(7):1887-1898. [doi: 10.1109/JBHI.2020.2980979] [Medline: 32175880]
67. Zia J, Kimball J, Inan O. Localizing placement of cardiomechanical sensors during dynamic periods via template matching. In: Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). 2020 Presented at: 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); July 20-24, 2020; Montreal, QC, Canada p. 473-476. [doi: 10.1109/EMBC44109.2020.9176732]
68. Guanqun Z, Da X, Olivier N, Mukkamala R. Pulse arrival time is not an adequate surrogate for pulse transit time in terms of tracking diastolic pressure. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2011 Presented at: Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Aug. 30 - Sept. 3, 2011; Boston, MA, USA p. 6462-6464. [doi: 10.1109/IEMBS.2011.6091595]
69. Shandhi MM, Semiz B, Hersek S, Goller N, Ayazi F, Inan OT. Performance analysis of gyroscope and accelerometer sensors for seismocardiography-based wearable pre-ejection period estimation. IEEE J Biomed Health Inform 2019 Nov;23(6):2365-2374 [FREE Full text] [doi: 10.1109/JBHI.2019.2895775] [Medline: 30703050]
70. De Ciuceis C, Porteri E, Rizzoni D, Corbellini C, La Boria E, Boari GE, et al. Effects of weight loss on structural and functional alterations of subcutaneous small arteries in obese patients. Hypertension 2011 Jul;58(1):29-36. [doi: 10.1161/HYPERTENSIONAHA.111.171082] [Medline: 21555677]

71. Ryan J, Edney S, Maher C. Anxious or empowered? A cross-sectional study exploring how wearable activity trackers make their owners feel. BMC Psychol 2019 Jul 03;7(1):42 [FREE Full text] [doi: 10.1186/s40359-019-0315-y] [Medline: 31269972]
72. Bove LA. Increasing patient engagement through the use of wearable technology. J Nurse Pract 2019 Sep;15(8):535-539. [doi: 10.1016/j.nurpra.2019.03.018]
73. Greiwe J, Nyenhuis SM. Wearable technology and how this can be implemented into clinical practice. Curr Allergy Asthma Rep 2020 Jun 06;20(8):36 [FREE Full text] [doi: 10.1007/s11882-020-00927-3] [Medline: 32506184]
74. Stergiou GS, Baibas NM, Gantzarou AP, Skeva II, Kalkana CB, Roussias LG, et al. Reproducibility of home, ambulatory, and clinic blood pressure: implications for the design of trials for the assessment of antihypertensive drug efficacy. Am J Hypertens 2002 Feb;15(2 Pt 1):101-104. [doi: 10.1016/s0895-7061(01)02324-x] [Medline: 11863243]
75. Lee K, Ni X, Lee JY, Arafa H, Pe DJ, Xu S, et al. Mechano-acoustic sensing of physiological processes and body motions via a soft wireless device placed at the suprasternal notch. Nat Biomed Eng 2020 Feb;4(2):148-158 [FREE Full text] [doi: 10.1038/s41551-019-0480-6] [Medline: 31768002]
76. Boutry CM, Beker L, Kaizawa Y, Vassos C, Tran H, Hinckley AC, et al. Biodegradable and flexible arterial-pulse sensor for the wireless monitoring of blood flow. Nat Biomed Eng 2019 Jan;3(1):47-57. [doi: 10.1038/s41551-018-0336-5] [Medline: 30932072]
77. Wang C, Li X, Hu H, Zhang L, Huang Z, Lin M, et al. Monitoring of the central blood pressure waveform via a conformal ultrasonic device. Nat Biomed Eng 2018 Sep;2(9):687-695 [FREE Full text] [doi: 10.1038/s41551-018-0287-x] [Medline: 30906648]
78. Ha T, Tran J, Liu S, Jang H, Jeong H, Mitbander R, et al. A chest-laminated ultrathin and stretchable e-tattoo for the measurement of electrocardiogram, seismocardiogram, and cardiac time intervals. Adv Sci (Weinh) 2019 Jul 17;6(14):1900290 [FREE Full text] [doi: 10.1002/advs.201900290] [Medline: 31380208]

Abbreviations

ABP: arterial blood pressure
AFE: analog front end
AO: aortic valve opening
BP: blood pressure
BPF: bandpass filter
DBP: diastolic blood pressure
ECG: electrocardiogram
IEEE: Institute for Electronics and Electrical Engineers
IR: infrared
LED: light-emitting diode
MAD: mean absolute difference
MAP: mean arterial pressure
ML: machine learning
MUA: medically underserved area
PCC: Pearson correlation coefficient
PEP: pre-ejection period
PPG: photoplethysmogram
PTT: pulse transit time
PWV: pulse wave velocity
RMSE: root mean square error
SBP: systolic blood pressure
SCG: seismocardiogram
SNR: signal-to-noise ratio

Edited by L Buis; submitted 26.01.21; peer-reviewed by J Rogers, S Omboni; comments to author 02.02.21; revised version received 10.02.21; accepted 10.05.21; published 02.08.21.

Please cite as:
Ganti V, Carek AM, Jung H, Srivatsa AV, Cherry D, Johnson LN, Inan OT
Enabling Wearable Pulse Transit Time-Based Blood Pressure Estimation for Medically Underserved Areas and Health Equity: Comprehensive Evaluation Study
JMIR Mhealth Uhealth 2021;9(8):e27466
URL: https://mhealth.jmir.org/2021/8/e27466
doi:10.2196/27466
PMID:34338646

Original Paper

General Practitioners' Perceptions of the Use of Wearable Electronic Health Monitoring Devices: Qualitative Analysis of Risks and Benefits

Lucia Volpato[1], MSc; María del Río Carral[1], PhD; Nicolas Senn[2], MD, PhD; Marie Santiago Delefosse[1], PhD

[1]Research Centre for Psychology of Health, Aging and Sport Examination, Institute of Psychology, University of Lausanne, Lausanne, Switzerland
[2]Department of Family Medicine, Centre for Primary Care and Public Health, University of Lausanne, Lausanne, Switzerland

Corresponding Author:
María del Río Carral, PhD
Research Centre for Psychology of Health, Aging and Sport Examination
Institute of Psychology
University of Lausanne
Géopolis Quartier UNIL - Mouline
Lausanne, 1015
Switzerland
Phone: 41 21 692 32 67
Email: maria.delriocarral@unil.ch

Abstract

Background: The rapid diffusion of wearable electronic health monitoring devices (wearable devices or wearables) among lay populations shows that self-tracking and self-monitoring are pervasively expanding, while influencing health-related practices. General practitioners are confronted with this phenomenon, since they often are the expert-voice that patients will seek.

Objective: This article aims to explore general practitioners' perceptions of the role of wearable devices in family medicine and of their benefits, risks, and challenges associated with their use. It also explores their perceptions of the future development of these devices.

Methods: Data were collected during a medical conference among 19 Swiss general practitioners through mind maps. Maps were first sketched at the conference and their content was later compared with notes and reports written during the conference, which allowed for further integration of information. This tool represents an innovative methodology in qualitative research that allows for time-efficient data collection and data analysis.

Results: Data analysis highlighted that wearable devices were described as user-friendly, adaptable devices that could enable performance monitoring and support medical research. Benefits included support for patients' empowerment and education, behavior change facilitation, better awareness of personal medical history and body functioning, efficient information transmission, and connection with the patient's medical network; however, general practitioners were concerned by a lack of scientific validation, lack of clarity over data protection, and the risk of stakeholder-associated financial interests. Other perceived risks included the promotion of an overly medicalized health culture and the risk of supporting patients' self-diagnosis and self-medication. General practitioners also feared increased pressure on their workload and a compromised doctor–patient relationship. Finally, they raised important questions that can guide wearables' future design and development, highlighting a need for general practitioners and medical professionals to be involved in the process.

Conclusions: Wearables play an increasingly central role in daily health-related practices, and general practitioners expressed a desire to become more involved in the development of such technologies. Described as useful information providers, wearables were generally positively perceived and did not seem to pose a threat to the doctor–patient relationship. However, general practitioners expressed their concern that wearables may fuel a self-monitoring logic, to the detriment of patients' autonomy and overall well-being. While wearables can contribute to health promotion, it is crucial to clarify the logic underpinning the design of such devices. Through the analysis of group discussions, this study contributes to the existing literature by presenting general practitioners' perceptions of wearable devices. This paper provides insight on general practitioners' perception to be considered in the context of product development and marketing.

(JMIR Mhealth Uhealth 2021;9(8):e23896) doi:10.2196/23896

KEYWORDS

mHealth; wearable devices; health wearables; activity trackers; health monitoring; self-tracking; general practitioners; mind maps; qualitative research; health psychology

Introduction

Over the last few years, the development of new health-related technologies has been particularly rapid and prosperous [1,2]. In particular, wearable electronic health monitoring devices (henceforth referred to as wearable devices or wearables) are designed to support health management by monitoring bodily vital signals, as well as tracking an individual's activity and habits [3,4]. User-friendliness and gamification play an important role in the appealing and engaging design of wearable devices, which are often paired with mobile phone apps [5]. Their association with mobile phones is at the origin of the terms *mHealth* (or *mobile health*) and continuous connection with wireless devices has been associated with self-surveillance and self-tracking mentality [6]. Yet, the difference between technologies targeting lay populations and the ones designed to monitor specific medical conditions is not always clearly defined [3]. In both fields—health promotion and intervention—the logic underpinning wearable devices' marketing aims to improve health by promoting behavior change through self-tracking, on the basis of feedback mechanisms [4,7,8].

More specifically, the marketing discourse on wearable devices is strongly based on the promise of benefits regarding personalized health management programs that claim to promote patient's self-responsibility and autonomy [6] by fostering a less hierarchical relationship between the user and the health professional [7]. Furthermore, wearable manufacturing companies have been sponsoring large-scale medical studies [9,10] with little consideration of the consequences associated with the introduction of self-tracking devices in everyday life [11]. This adds further complexity to the picture because traditional health care systems risk being detoured by other—often profit-led—motives [11]. While it has been argued that wearable devices may contribute to fostering users' autonomy [12], self-tracking has not always been associated with empowerment [13,14]. In fact, wearable use could hinder the autonomy of users, who would increasingly rely on these devices for daily, health-related decision making [4]. According to Andreassen and colleagues [15], feelings of domestication and resistance co-exist in the user–object relationship. Furthermore, wearable devices have been associated with high abandonment rates after only 6 months of use [16]. It is, therefore, clear that, beyond manufacturers' promises, the concrete use of these technologies in the health sector is subject to negotiation processes that depend on complex dynamics regarding the object–user relationship and the doctor–patient relationship [17,18].

Activity trackers have affected how treatment, visits, and health management are established during health consultations [19]. Their growing presence may constitute a positive addition to the relationship between patients and their general practitioners (GPs), because the devices could support effective transmission of health-related information [19]. Historically, the introduction of technological devices has defined medical practices [20], and wearable devices are no exception. The notion of the Quantified Self—associated with activity trackers—is pervasively shaping health norms through self-surveillance across life domains [21]. Moreover, the contemporary trend of healthism, which values individual responsibility and surveillance in health management, has expanded over the past decades [22]. This societal discourse constitutes a fertile ground for the production and marketing of wearables.

GPs are an inherent part of the digital health revolution, given their role as health experts, citizens, and sometimes, wearable consumers [23]. The intention to adopt activity trackers and recommend their use are shaped by views and attitudes that individuals may have toward such technologies [24,25]. In light of these findings, it is necessary to further explore perception and views on wearables among GPs, because such devices may change existing medical practices and contribute to shape new ones. This is particularly relevant within the Swiss context, where, similarly to other European countries [26,27], GPs are in the front-line when patients access the health system. In this sense, often GPs are the first health professionals to interact with many wearable device users.

Some studies have investigated health professionals' attitudes toward technologies that are specific to chronic health conditions, such as epilepsy [28], asthma [29], arthritis [30], and other chronic diseases [31]. Another body of literature on GPs' perspectives has examined their experience with a wide array of eHealth innovations, beyond the specific use of wearables [32]. With respect to wearables, few studies have explored GPs' attitudes with a set of predefined topics using individual semistructured interviews [33,34] and web-based surveys [19]. We further investigated GPs' perspectives and contribute to the existing literature on the role of wearables in family medicine. We aimed to explore how GPs perceive wearable devices—both for health promotion and clinical use—in the context of their medical practice, by focusing on perceived benefits and risks. To do so, we used an innovative qualitative methodology with mind maps to analyze group discussions that took place during a medical conference on family medicine. Mind maps have been described as being particularly suitable for analyzing group discussions in the field of health care [35]. We discuss salient elements to consider in the future development of these technologies.

Methods

Research Context and Sampling

This study's aims were defined by the authors in collaboration with health psychologists and physicians working at the University of Lausanne, as well as GPs working in the French-speaking regions of Switzerland. Data were collected in a symposium—New Technologies in Family Medicine—that took place in Switzerland in 2018, as part of a medical conference, mainly targeting GPs. Given the qualitative nature

of our study, we aimed at in-depth understanding and contextualization of data, rather than generalization. For data collection and analysis, we followed the quality criteria for qualitative research defined in the field of health psychology [36,37].

We used convenience sampling: 19 GPs (7 female and 12 male) working in family medicine in the French-speaking regions of Switzerland. Participants were informed about the symposium's goal, involving the definition of potential research perspectives regarding the use of wearables in family medicine, based on their perception. GPs were formally informed that group discussions would be recorded for further analysis, and oral consent was obtained. Under Swiss ethical regulations, no written consent was required as no biomedical information was collected.

Regarding participants' background on wearables use, the vast majority reported not having actively introduced them in their clinical practice and that any discussion on wearables was usually initiated by patients themselves. Cited examples included patients who monitored their menstrual cycle through apps and the tracking of physical activity through smartwatches. Only 2 GPs reported that they used mobile apps for sleep monitoring and for diagnostic procedures via symptom-input mechanisms. It was highlighted that such apps were offered by official health providers. Some participants were familiar with such technologies through life experiences beyond their professional practice as GPs. With respect to their personal use, 1 GP reported using a smartwatch for performance monitoring during sports training. In contrast, another participant reported deactivating all tracking functions on their mobile device because of mistrust of the app's use of personal information.

Group Discussions

GPs were enrolled in group discussions on smartwatches, wearable devices, and health apps. These topics had been previously defined, so that participants could join any group discussion, based on their personal interests. Each group (average of 6 participants per group) was moderated by 1 health psychologist and 1 GP. Discussions lasted approximately 1 hour and were audiorecorded.

In each group, participants were invited to briefly present themselves and were informed that the discussions were going to address the role of multiple mHealth technologies in family medicine. The 2 moderators introduced a brief explanation of the specific discussion topic (either smartwatches, wearable devices, or health apps, each discussed within a different group). These 3 groups of technologies were chosen for their high interconnectedness and interdependence within the broad category of mHealth [6]. For instance, smartwatches may be considered a wearable device category and are often supported by a smartphone app for data collection and analysis [5]. Participants were asked to discuss the following questions within each group: (1) What is the role of such technologies within your practice, according to your experience? (2) What risks and benefits do you identify in relation to such technologies? (3) Which challenges would you associate to the concrete use in your professional practice?

While the differences with other methods of data collection (ie, group interviews or focus groups) may be subtle, group discussions are less bound to structured interview guides and the emerging discussion topics often result from the interactions among group members, rather than from detailed predetermined questions grid [38]. Group discussions are also particularly suited for data collection among individuals who belong to the same group, for example, a professional category [38,39]. Moreover, the role of the moderators in group discussions is to provide topics to stimulate interactions among participants in a nondirective way [38].

To facilitate participants' interactions, moderators took part in the discussions and summarized the material produced from their group. Summaries were approved and validated by participants of each group, resulting in specific descriptive reports [40-42]. Participant validation has shown to be a critical stage of qualitative research, because it provides more solidity and pertinence to the collected data [37].

Mind Maps

The potential of mind maps has been recently underlined for their use as research methods for data collection and analysis in the field of health [35]. A mind map can be defined as "a diagram used to represent concepts, ideas or tasks linked to and arranged radially around a central key word or idea [35]." Mind maps present information in a hierarchical way [43] through a synthetic visually engaging format [44]. Beyond their use in data collection, they can facilitate the data analysis by identifying and representing thematic and conceptual patterns in a nonlinear form [45], while showing associations between ideas and topics [43]. During the symposium, an overt participant–observant researcher (the main author) circulated among the different discussion groups, taking notes of part of the ongoing discussions, and sketching preliminary mind maps. Through participant observation, further notes were taken to identify the links between the raised concepts and capture the contextual dimensions of verbal exchanges [46].

Inspired by Burgess-Allen and Owen-Smith [35], we considered a separate mind map for each of the themes used for the 3 groups canvases: benefits, risks, and some insights for the future. In addition, mind maps drawn during the group discussions revealed a recurring substructure in the discussion of the themes: doctor–patient relationship, patient–device relationship and GPs' broad concerns, and final mind maps reflected this structure. The content of each mind map was then assembled inductively, and narrative contents were systematically compared by assessing their semantic similarities and differences. The proceedings and the audiorecordings from the group discussions facilitated integration of any missing relevant information from the preliminary notes and mind map drafting. This was particularly helpful to confirm the accuracy of the qualitative material and the mind map analysis. Finally, the 3 mind maps were compared to one another to identify common issues raised across group discussions regarding the potential benefits and risks, as well as, some insights for the future of wearable devices. This technique allowed for data analysis according to a theme-categories-subcategories structure, analogous to inductive thematic analysis, where mind mapping

is a preliminary stage [47]. In this study, mind maps were first sketched on paper for conceptualization purposes and were later digitally reproduced with FreeMind software (version 1.0.1).

Results

General

Regarding GPs' perceptions of smartwatches, wearable devices, and health apps in family medicine, the first mind map (Figure 1) summarizes the perceived benefits of wearable devices. Here, participants used the conditional verb tense, which suggested that their arguments often applied to hypothetical scenarios and specific conditions. The second map (Figure 2) shows perceived risks that wearable devices usage and promotion may entail. The third map (Figure 3) presents insights that should be considered in the future production and use of wearables.

Figure 1. Mind map presenting the benefits that general practitioners associate with wearable use.

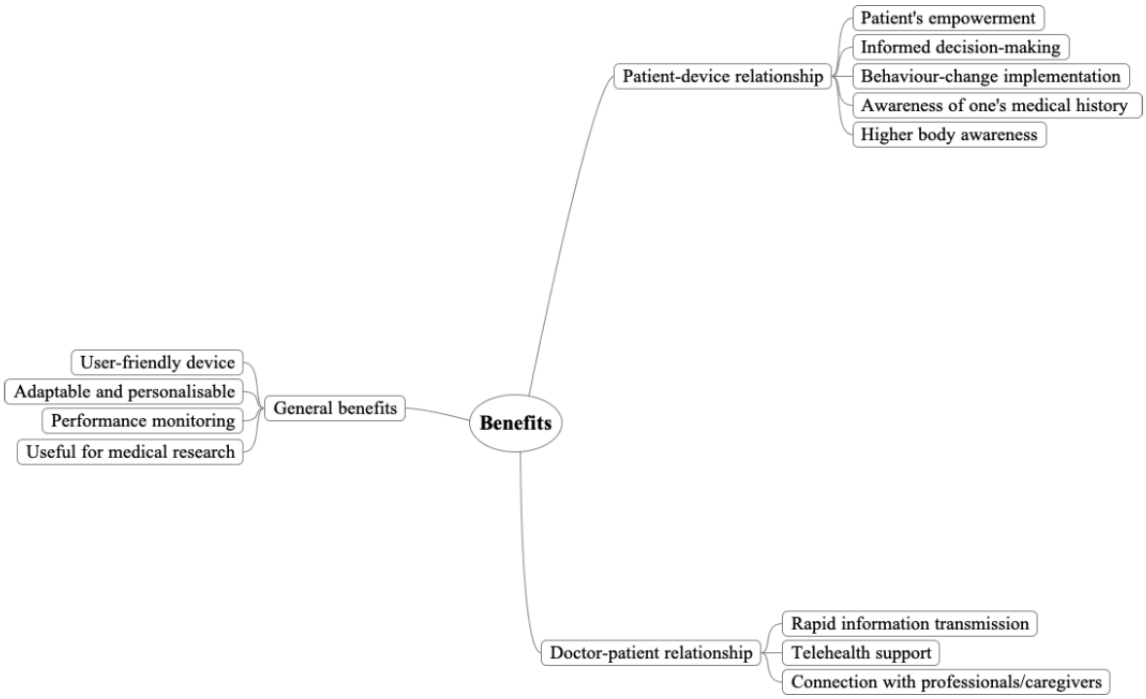


Figure 2. Mind map presenting the risks that general practitioners associate with wearable use.

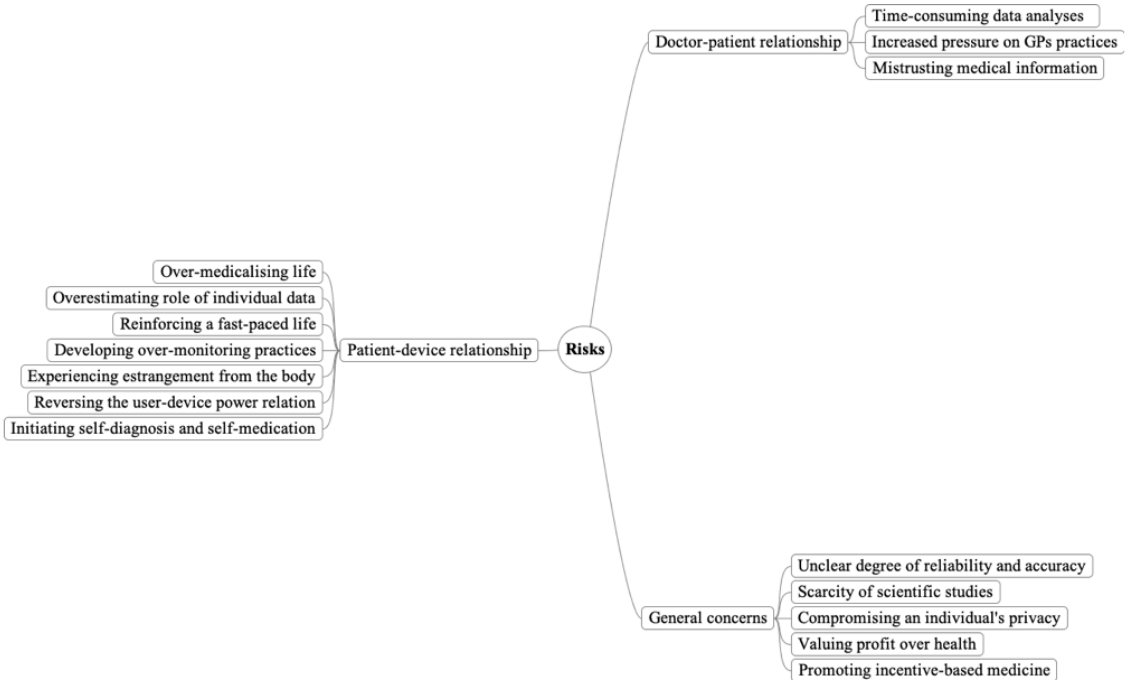
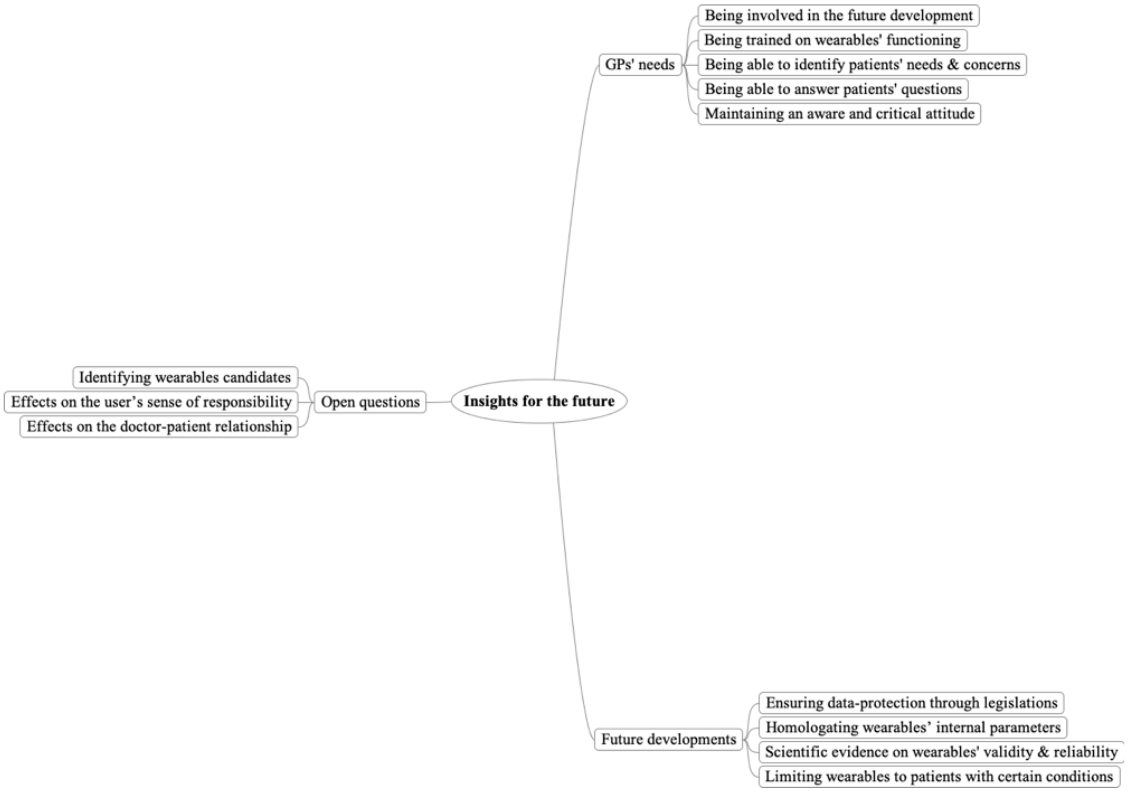


Figure 3. Mind map presenting insights for the future of wearables according to general practitioners.



Benefits That GPs Associate With Wearable Use

General Benefits Associated With Wearable Devices' Characteristics

Overall, GPs positively evaluated wearable devices that were considered user-friendly to be used in a variety of health situations, thereby representing an attractive solution for different populations. According to GPs, wearable devices could be easily used by both patients and health professionals due to their simple, intuitive designs. In particular, doctors appreciated devices whose parameters can be easily adapted and personalized to fit a patient's personal health characteristics. In fact, these features could enable a personalized approach to health management. Moreover, GPs considered that wearables could benefit those who wished to have regular feedback about their personal health and measure their physical activity through performance monitoring. According to GPs from the 3 group discussions, the widespread use of wearables could allow for large-sample data collection, which would be especially useful for medical research. In fact, a potential benefit concerned the strong statistical power that wearable devices could enable through research conducted among a high number of users.

Patient–Wearable Relationship

Wearable devices were also discussed in relation to patients' empowerment. According to different participants across the 3 group discussions, wearables could help raise awareness among patients on their overall health condition. Due to the feedback and reward mechanisms that define certain devices, wearables could train users to make informed health decisions. In this sense, results reveal that wearables may enhance patient's self-responsibility and be a concrete partner for health promotion. In fact, wearables were described as a potential mean for behavior change through the implementation of new health behaviors, through consistent self-management. As stated by a participant:

> *Smart watches could motivate people for being more active, because there's a certain degree of satisfaction in seeing the [step] counting going up.*

Moreover, some GPs asserted that wearables could help patients to keep track of their medical history and develop an unprecedented awareness regarding their own bodies. For instance, wearable users could establish links between their own feelings and the data provided by the device. In their views, wearables may, in fact, concretely support self-education.

GP–Patient Relationship

Wearables were described as potential health care partners in the light of the rapid information-transmission processes that such devices enable, allowing patients to have a more central role in their health management. Furthermore, the time-efficient data exchange between patients and health professionals could be beneficial for the management of certain health conditions, such as epilepsy or cardiac diseases. These health conditions were mentioned given the capacity of wearables to transfer data in real time. From GPs' views, this represented a helpful feature to prevent seizures or enable screening procedures. Wearables were also perceived as potential allies for telehealth, since these devices could help GPs reach patients living in geographically remote areas.

One GP affirmed:

> *We don't need to make a trip to their place if they can measure themselves blood tension, glycemia, the*

> *heartrate at home and then transfer information via the internet.*

This was considered a particularly important aspect in the Swiss context, where, due to alpine geography, certain patients may need to travel long distances to receive medical care. Wearables were described as a possible means of connection within the health system between different professionals and caregivers, as well as a useful solution for those who do not have family or social support in daily health care.

Risks That GPs Associate With Wearable Use

General Risks Associated With Wearable Devices' Characteristics

Although wearables' potential advantages were thoroughly discussed, the debate raised several potential risks linked to their use. For instance, participants expressed fears regarding the unclear degree of reliability and accuracy of commercial wearables that are increasingly available. GPs shared their professional experience in stressing that devices that are able to accurately record biometrics are often more expensive and more complex. Therefore, their use requires specific training and an understanding of data collection. Furthermore, GPs highlighted the scarcity of scientific studies on wearables' validity and reliability. To them, this represented an obstacle that impeded them from actively promoting the use of instruments that are not supported by scientific evidence.

Furthermore, the lack of accurate information regarding the management of biomedical data by manufacturers was considered to be a serious danger for patients. The risk of compromising an individual's privacy was a major concern with respect to these technologies since, as affirmed by a participant:

> *Third-party use of personal data is still very poorly regulated.*

In an era where personal data is becoming widely commodified, several industries can profit from wearable use without being genuinely concerned by users' health. According to GPs, the promotion of wearables could thus imply that financial profit is valued over health.

The role of health insurance companies and their possible relation with the wearable device industry were also considered in group discussions, since the Swiss health system relies on compulsory private not-for-profit health insurance companies. According to some participants, the collaboration between the 2 stakeholders could encourage the development of incentive-based medicine rationalities, that is, of a health philosophy, by which patient behavior could be rewarded or punished by insurance companies as a consequence of the degree of behavior compliance determined by the wearable's design. For instance, health insurance companies could be inclined to reward so-called good users for having achieved the health aims set by the device or punish so-called bad users who have failed to do so. This mechanism deserves better attention, because it may also potentially reinforce health inequalities from a socioeconomic perspective. These reflections raised a debate on which ethical principles should underpin family medicine, as well as, on the rights and responsibilities of each actor.

Patient–Wearable Relationship

Regarding user–device relationship, GPs argued that continuous self-monitoring could stress an overmedicalization of life, generated by an excessive intellectualization of the user's physical condition. From their perspective, the prioritization of self-monitoring practices in the field of health would inevitably confront users with the paradoxes of our culture: while health-related practices are aimed at reducing stress in daily life by helping users to slow their pace, wearables would be the symbol of a society that values rapid information exchange, and hence, would contribute to reinforcing a fast-paced life. In this context, GPs raised the risk of overestimating the value and the role of individual data in coming to conclusions about a person's general state of health. According to GPs, the continuous measurement of biomedical information appeared to be also potentially anxiety triggering. A participant feared that

> *People may end up spending more time preoccupying about their health instead of living.*

Patients with apprehensive personalities could particularly risk developing overmonitoring practices, to the detriment of their mental health.

Regarding the level of trust toward certain devices, some participants feared that wearables would induce the users to gradually feel estranged from their body. In this sense, wearables could provide digital information that does not correspond to the users' subjective perceptions on their own body and health. This mismatch between the wearable's feedback and the embodied sensations could induce the users to mistrust their subjective perception and thus feel disconnected from their own body. According to GPs, this risk would also interfere with the principle of patient autonomy, whose appraisal of their own body would therefore mainly depend on the wearable verdict instead of their own perception. In this scenario, the patient and the caregiver would need to invest even more resources to set up a process of bodily re-appropriation. From the participants' view, these risks would result in a reversed power relationship with the device that could be dangerous and that should be avoided. A participant feared a

> *Very likely tendency towards over-training during a sport session*

while seeking positive feedback from an activity tracker. An important element of the debate concerned wearable data interpretation. Participants agreed on the fact that, given a decrease of exchanges between users and health professionals, the former would be confronted to increased uncertainties regarding the interpretation of their personal physiological values, which is considered to be as dangerous for a user's health. In fact, on the basis of the wearable data, patients could be tempted with self-diagnosis or self-medication solutions, something that ought to be avoided, especially when medical expertise is essential.

GP–Patient Relationship

In the light of the increased production of patient-specific medical information, participants highlighted the risk of devoting their working hours to time-consuming data analyses.

> *We can collect plenty of data, but then what will do about them?*

wondered a participant. In fact, the instantaneous nature data transmission could intensify the expectation of an immediate reply from health professionals, which would amplify the pressure on GPs' daily practices, without any verified benefit for concerned patients. Participants also expressed the danger of inconsistency between the information recorded by wearables and data provided by other devices measuring biometrics. This type of divergence could, in fact, entail a progressive mistrust on the part of patients regarding the information provided by other the medical instruments, and GPs may suffer from credibility loss.

Insights for the Future of Wearables According to GPs

GPs' Professional Needs

Participants expressed a sense of inevitability toward the introduction of wearable devices into contemporary medical practices within the Swiss context, regardless of the outcomes of current research in the field. In this sense, several GPs highlighted the urgent need, and their personal interest, to become more involved in the development of wearables. In their view, a synergy between producers and health professionals is necessary to enable the design of beneficial instruments. GPs also expressed the wish to receive training to better understand how these technologies work (especially concerning data collection and storage of information) and to be better informed on the news regarding the health-wearable market. GPs particularly valued the importance of understanding patients' concerns, of identifying patients' health needs, and of answering to their questions.

> *When worried patients come [to the consultation] with such instruments, what do we do? We have to give them an answer,*

stated a participant. They also expressed the importance to them as professionals to be aware and to remain critical of the usefulness these devices in health care.

Future Developments

Overall, participants from the 3 group discussions raised the urgent need for firm legislation to guide future design, production, and marketing of wearables. In particular, they showed a high degree of mistrust regarding the confidentiality of their patients' data that wearable companies could guarantee. To them, a priority for the future would therefore, be legislation to ensure data protection, as well as an overt policy to safeguard the rights of users, because the latter are in a vulnerable position within the health care system. Furthermore, the legislation should also concern the homologation of wearables' internal parameters. Alongside scientific research, these measures could help to produce more valid and reliable devices for health self-management. With respect to the role that wearables should have, participants expressed the vision of wearables as partners that could help improve the care of patients. As emphasized by certain participants, well-oriented and accurate feedback is central to medical practice, because it can facilitate the learning process. In this sense, the analyses showed that wearables would not only be a tool for information transmission but could solidify partners in the promotion of behavior change. Moreover, according to the participants, the use of wearables should be limited to patients suffering from certain health conditions (although no examples were explicitly given during the symposium), instead of monitoring healthy people. In this sense, health would be achieved by conducting a digital-free, slow-paced life, where the person is not dependent on self-tracking devices.

Open Questions

According to the participants, several questions remained unanswered yet would be worth exploring. For instance, some GPs wondered how to assess which patients would best benefit from wearable use, and on the contrary, which patients would feel disconnected from their own body, that is, lose their personal autonomy toward the interpretation of their own embodied feelings. In this sense, a participant asserted that

> *To be useful, such products could be adapted to the patient's profile in the future.*

GPs also wished to know how wearables may affect user's sense of responsibility. More generally GPs also raised the following issue: How will health wearables affect the GP–patient relationship? These questions closed the debates across groups, highlighting the need of additional analyses before establishing any further statements on the role of health wearables for our contemporary societies.

Discussion

Principal Findings and Comparison With Prior Work

GPs play a crucial role in the health care system by promoting and prescribing specific health practices. We aimed to explore their perceptions on wearables in the context of family medicine, by directly addressing the risks and benefits associated with these technologies and reflecting on possible future developments in the field of health care. The methodology adopted in this research was qualitative. This perspective was particularly suited to generate exploratory and contextualized knowledge. While group discussions allowed us to capture GPs' views throughout their spontaneous interactions, mind maps enabled an iterative and efficient process of data collection and analysis given our research setting [35].

Wearables as Information Providers

The effects of digital health wearables on the doctor–patient relationship appeared to be both beneficial and risky, highlighting their ambiguous potential. While wearables were viewed as suitable for information transmission, coordination, and general illness management, GPs also feared that these technologies would put increased pressure on their role and expertise as health professionals. Indeed, GPs anticipated longer consultations that would be dedicated to data analysis and data interpretation stemming from patients' wearables. From this perspective, our results confirm those of recent studies showing that wearables are considered to be particularly useful for information-transmission and general illness management but that time-consuming data interpretation continue to be important concerns among health professionals [19,34,48]. With respect

to digital information, GPs also expressed their concern regarding product reliability and patients' data protection. As recent studies have argued, developers need to consider these key issues when designing health-monitoring technologies [19,48,49].

Participants perceived wearables as user-friendly devices that could foster patients' empowerment and support them throughout behavior change processes [33,34]. However, the use of wearables for patient education and empowerment has also been associated with a patronizing view of the doctor–patient relationship [6]. Preventing such repercussions represents a concrete challenge faced by research in the health sector. In this context, wearable use can be envisaged in relation to the concept of continuity in health care, defined as informational, relational, or management related [50]. In this sense, wearables constitute tools that can positively contribute to ensuring informational and health management continuity. Nevertheless, these tools alone may not be able to support the multifaceted relationship continuity between the doctor and the patient and would hence need to be adapted.

Self-tracking: A Catalyst for Healthism

GPs were also concerned about the role that wearables could play in patients' everyday lives outside medical consultations. For instance, GPs highlighted the potential risk of promoting a dominant social discourse or life-philosophy, where self-tracking and self-monitoring become practices that are encouraged, even among individuals who are healthy or who do not suffer from specific health conditions. This overmedicalization of life can be compared to what Gabriels and colleagues [33] have coined as *entertainment medicine*, where self-tracking devices become responsible for producing "medically unnecessary data that belong more to the fitness or wellness than to the medical realm." Echoing past literature [32], GPs stressed the importance of understanding patients' needs in order to address their concerns more effectively.

More generally, self-tracking in the medical field has been previously argued to have culturally and structurally transformed the ways in which health-related practices are being defined [15]. In this sense, an important contribution of our findings to the debate is GPs' strong resistance to incentive-based medicine, in which healthy behaviors are implemented within a reward versus punishment mechanism. This posture contests 2 aspects. The first is with respect to the global trend across stakeholders to collect information produced by wearable devices for financial purposes [51], which causes ethical concerns to be raised by GPs. The second refers to the philosophical and pedagogical premises underpinning incentive-based medicine. In GPs' views, this type of medicine contrasts with the value of patient autonomy and risks to promoting an undesirable obsessive compliance with health standards set by wearables. In the contemporary dominant culture of healthism that values self-management [22], this risk becomes increasingly important. Through subtle imperatives, wearables may indeed respond to patients' intention to take control over their own health [21], while simultaneously triggering feelings of apprehension and self-inadequacy. GPs' intentions of promoting patient autonomy emphasizes the urgent need to develop alternative approaches in health care that can facilitate behavior change. Indeed, as in the case of other social practices, health practices are subject to ambiguity, contradictions, and ultimately, continuous change [18]. We argue that these premises should be considered in the design of wearable technologies.

Future Perspectives

The rapid expansion of wearables has entailed changes that remain unchallenged regarding their social, psychological and cultural implications for individual and public perceptions of health within our Western societies dominated by healthism [22]. In this sense, it is essential to clarify the rationale underpinning the development and marketing of such devices, whose extensive use may not necessarily be desirable from a GP's perspective. A clear legal frame guiding the production and distribution of wearables for medical usage might help guide the effectiveness and clinical safety for users and health professionals. For instance, the concept of Health Technology Assessment [52] offers a useful illustration of how this frame could be conceptualized. This study calls for future research to deconstruct and analyze the logic behind the conceptualization, development, and use of health wearables, from the perspective of health professionals, users, and technology developers. In this context, it would be interesting to compare these results with patients' views, in order to identify possible differences, with an aim toward better integration of wearables in general medical practice. Indeed, our study confirms the necessity for researchers and developers to question the values and logic guiding wearable design.

Limitations

This study is not exempt from limitations. Given its exploratory nature, our qualitative results require further analysis regarding other contexts and methodologies. Moreover, while appropriate to our research setting, mind maps allow limited in-depth data analysis compared to other qualitative methods [35]. In addition, the visually synthetic characteristic of mind maps does not allow for data saturation claims and does not allow the integration of specific details. Rather, mind maps constitute an exploratory step in research that can be complemented by other techniques [47]. Nonetheless, this method is useful to develop hypotheses that can be tested in future research.

Conclusion

This study found that GPs are willing to be more actively engaged as collaborators in the design, development, and promotion of wearables, alongside producers and end-users. Our research contributes to broadening current understanding of wearables and self-tracking technologies in the field of family medicine, by emphasizing the role of wearables as key information providers. Indeed, GPs are neither passive spectators of—nor opponents to—digital health developments, which are perceived to be increasingly more important and inevitable. In spite of the important role of wearables, this study underlined the irreplaceable character of the doctor–patient relationship, which remains a central dimension in family medicine. GPs manifested their opposition to the logic of self-monitoring that GPs considered to have a negative impact on patients' global well-being and autonomy. Regarding research perspectives, it

seems crucial to reflect upon the definition of health that is being shaped by wearables and similar self-tracking technologies. These perspectives would enable an informed comparison across main actors in health care and contribute to collective coordinated efforts to improve individual and public health while reducing health-related costs.

Acknowledgments

We would like to thank the general practitioners who participated to the workshops. We would also like to thank the moderators: Vincent Gesbert, PhD; Roane Keller, MD; Thierry Mathieu; Daniel Widmer, MD; Baptiste Pedrazzini, MD; and Vanlisa Bourqui. We would like to express our gratitude toward *Les Editions Médecine et Hygiène* for their kind support in the organization of the symposium.

Conflicts of Interest

None declared.

References

1. Lupton D. Quantifying the body: monitoring and measuring health in the age of mHealth technologies. Critical Public Health 2013 Dec;23(4):393-403. [doi: 10.1080/09581596.2013.794931]
2. Yetisen AK, Martinez-Hurtado JL, Ünal B, Khademhosseini A, Butt H. Wearables in medicine. Adv Mater 2018 Jun 11:e1706910 [FREE Full text] [doi: 10.1002/adma.201706910] [Medline: 29893068]
3. Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: promises and barriers. PLoS Med 2016 Feb;13(2):e1001953 [FREE Full text] [doi: 10.1371/journal.pmed.1001953] [Medline: 26836780]
4. 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]
5. Altenhoff B, Vaigneur H, Caine K. One step forward, two steps back: the key to wearables in the field is the app. 2015 Presented at: 9th International Conference on Pervasive Computing Technologies for Health care (PervasiveHealth); May 20-23; Istanbul, Turkey p. 241-244. [doi: 10.4108/icst.pervasivehealth.2015.259049]
6. Lupton D. M-health and health promotion: the digital cyborg and surveillance society. Soc Theory Health 2012 Jun 27;10(3):229-244. [doi: 10.1057/sth.2012.6]
7. 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]
8. Williamson B. Algorithmic skin: health-tracking technologies, personal analytics and the biopedagogies of digitized health and physical education. Sport Educ Soc 2014 Oct 26;20(1):133-151. [doi: 10.1080/13573322.2014.962494]
9. Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, for the Apple Heart Study Investigators. Large-scale assessment of a smartwatch to identify atrial fibrillation. N Engl J Med 2019 Nov 14;381(20):1909-1917. [doi: 10.1056/NEJMoa1901183] [Medline: 31722151]
10. Turakhia MP, Desai M, Hedlin H, Rajmane A, Talati N, Ferris T, et al. Rationale and design of a large-scale, app-based study to identify cardiac arrhythmias using a smartwatch: the Apple heart study. Am Heart J 2019 Jan;207:66-75 [FREE Full text] [doi: 10.1016/j.ahj.2018.09.002] [Medline: 30392584]
11. Campion EW, Jarcho JA. Watched by Apple. N Engl J Med 2019 Nov 14;381(20):1964-1965. [doi: 10.1056/NEJMe1913980] [Medline: 31722159]
12. Bradway M, Årsand E, Grøttland A. Mobile Health: empowering patients and driving change. Trends Endocrinol Metab 2015 Mar;26(3):114-117. [doi: 10.1016/j.tem.2015.01.001] [Medline: 25727462]
13. Nelson EC, Verhagen T, Noordzij ML. Health empowerment through activity trackers: an empirical smart wristband study. Comput Human Behav 2016 Sep;62:364-374. [doi: 10.1016/j.chb.2016.03.065]
14. Ryan J, Edney S, Maher C. Anxious or empowered? a cross-sectional study exploring how wearable activity trackers make their owners feel. BMC Psychol 2019 Jul 03;7(1):42 [FREE Full text] [doi: 10.1186/s40359-019-0315-y] [Medline: 31269972]
15. Andreassen HK, Dyb K, May CR, Pope CJ, Warth LL. Digitized patient–provider interaction: how does it matter? a qualitative meta-synthesis. Soc Sci Med 2018 Oct;215:36-44. [doi: 10.1016/j.socscimed.2018.08.036]
16. Ledger D, McCaffrey D. Inside wearables: how the science of human behavior change offers the secret to long-term engagement. Endeavour Partners. 2014. URL: https://medium.com/@endeavourprtnrs/inside-wearable-how-the-science-of-human-behavior-change-offers-the-secret-to-long-term-engagement-a15b3c7d4cf3 [accessed 2021-07-28]
17. Nettleton S, Burrows R. E-scaped medicine? information, reflexivity and health. Crit Soc Policy 2003 May 01;23(2):165-185. [doi: 10.1177/0261018303023002003]

18. Piras EM, Miele F. Clinical self-tracking and monitoring technologies: negotiations in the ICT-mediated patient–provider relationship. Health Sociol Rev 2016 Aug 05;26(1):38-53. [doi: 10.1080/14461242.2016.1212316]
19. Loos JR, Davidson EJ. Wearable health monitors and physician-patient communication: the physician's perspective. In: Proceedings of the 49th Hawaii International Conference on System Sciences. 2016 Presented at: 49th Hawaii International Conference on System Sciences; Jan 5-8; Koloa, HI, USA p. 3389-3399. [doi: 10.1109/hicss.2016.422]
20. Reiser SJ. Medicine and the Reign of Technology. Cambridge: Cambridge University Press; 1981.
21. Dagiral E, Licoppe C, Martin O, Pharabod AS. Le Quantified Self en question. Réseaux 2019;4:17-54. [doi: 10.3917/res.216.0017]
22. Crawford R. Healthism and the medicalization of everyday life. Int J Health Serv 1980;10(3):365-388. [doi: 10.2190/3H2H-3XJN-3KAY-G9NY] [Medline: 7419309]
23. 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]
24. 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 Sep 24;7(1):8099 [FREE Full text] [doi: 10.4081/hpr.2019.8099] [Medline: 31583292]
25. Talukder MS, Chiong R, Bao Y, Hayat Malik B. Acceptance and use predictors of fitness wearable technology and intention to recommend. IMDS 2019 Feb 04;119(1):170-188. [doi: 10.1108/imds-01-2018-0009]
26. Blinkenberg J, Pahlavanyali S, Hetlevik Ø, Sandvik H, Hunskaar S. General practitioners' and out-of-hours doctors' role as gatekeeper in emergency admissions to somatic hospitals in Norway: registry-based observational study. BMC Health Serv Res 2019 Aug 14;19(1):568 [FREE Full text] [doi: 10.1186/s12913-019-4419-0] [Medline: 31412931]
27. Damoiseaux RAMJ. [E-health and the GP as gatekeeper; are the two compatible?]. Ned Tijdschr Geneeskd 2014 Dec 18;158:A8558. [Medline: 25515393]
28. Bruno E, Simblett S, Lang A, Biondi A, Odoi C, Schulze-Bonhage A, RADAR-CNS Consortium. Wearable technology in epilepsy: the views of patients, caregivers, and healthcare professionals. Epilepsy Behav 2018 Aug;85:141-149. [doi: 10.1016/j.yebeh.2018.05.044] [Medline: 29940377]
29. Himes BE, Leszinsky L, Walsh R, Hepner H, Wu AC. Mobile health and inhaler-based monitoring devices for asthma management. J Allergy Clin Immunol Pract 2019 Nov;7(8):2535-2543 [FREE Full text] [doi: 10.1016/j.jaip.2019.08.034] [Medline: 31706485]
30. Leese J, Macdonald GG, Tran BC, Wong R, Backman CL, Townsend AF, et al. Using physical activity trackers in arthritis self-management: a qualitative study of patient and rehabilitation professional perspectives. Arthritis Care Res (Hoboken) 2019 Feb 29;71(2):227-236 [FREE Full text] [doi: 10.1002/acr.23780] [Medline: 30295430]
31. 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]
32. Brandt CJ, Søgaard GI, Clemensen J, Sndergaard J, Nielsen JB. General practitioners' perspective on ehealth and lifestyle change: qualitative interview study. JMIR Mhealth Uhealth 2018 Apr 17;6(4):e88 [FREE Full text] [doi: 10.2196/mhealth.8988] [Medline: 29666045]
33. 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]
34. Watt A, Swainston K, Wilson G. Health professionals' attitudes to patients' use of wearable technology. Digit Health 2019;5:2055207619845544 [FREE Full text] [doi: 10.1177/2055207619845544] [Medline: 31041111]
35. Burgess-Allen J, Owen-Smith V. Using mind mapping techniques for rapid qualitative data analysis in public participation processes. Health Expect 2010 Dec;13(4):406-415 [FREE Full text] [doi: 10.1111/j.1369-7625.2010.00594.x] [Medline: 20550595]
36. Santiago-Delefosse M, Gavin A, Bruchez C, Roux P, Stephen SL. Quality of qualitative research in the health sciences: analysis of the common criteria present in 58 assessment guidelines by expert users. Soc Sci Med 2016 Jan;148:142-151. [doi: 10.1016/j.socscimed.2015.11.007] [Medline: 26698197]
37. Willig C. Introducing Qualitative Research in Psychology 3rd ed. Maidenhead, Berkshire, United Kingdom: Open University Press; 2013.
38. Flick U. An Introduction to Qualitative Research 6th ed. London: Sage; 2018.
39. Bohnsack R. Group discussion and focus groups. In: Flick U, von Kardoff E, Steinke I, editors. A Companion to Qualitative Research. London: Sage; 2004:214-221.
40. Gesbert V, Keller R. Montres connectées et pratique médicale. Revue Médicale Suisse 2018 Dec 05;14(S21-S29):11-12 [FREE Full text]
41. Mathieu T, Widmer D. Quels défis pour la pratique médicale à l'ère des applications de santé? Revue Médicale Suisse 2018 Dec 05;14(S21-S29):17-19 [FREE Full text]

42. Pedrazzini B, Bourqui V. Place de l'objet de santé connecté en médecine de famille. Revue Médicale Suisse 2018 Dec 05;14(S21-S29):13-14 [FREE Full text]
43. Davies M. Concept mapping, mind mapping and argument mapping: what are the differences and do they matter? High Educ 2010 Nov 27;62(3):279-301. [doi: 10.1007/s10734-010-9387-6]
44. Eppler MJ. A comparison between concept maps, mind maps, conceptual diagrams, and visual metaphors as complementary tools for knowledge construction and sharing. Inf Vis 2006 Jun 22;5(3):202-210. [doi: 10.1057/palgrave.ivs.9500131]
45. Biktimirov E, Nilson L. Show them the money: using mind mapping in the introductory finance course. J Fin Educ 2006;32:72-86 [FREE Full text]
46. Aagaard J, Matthiesen N. Methods of materiality: participant observation and qualitative research in psychology. Qual Res Psychol 2015 Oct 07;13(1):33-46. [doi: 10.1080/14780887.2015.1090510]
47. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
48. Nasir S, Yurder Y. Consumers' and physicians' perceptions about high tech wearable health products. Procedia Soc Behav Sci 2015 Jul;195:1261-1267. [doi: 10.1016/j.sbspro.2015.06.279]
49. Bove LA. Increasing patient engagement through the use of wearable technology. J Nurse Pract 2019 Sep;15(8):535-539. [doi: 10.1016/j.nurpra.2019.03.018]
50. Haggerty JL, Reid RJ, Freeman GK, Starfield BH, Adair CE, McKendry R. Continuity of care: a multidisciplinary review. BMJ 2003 Nov 22;327(7425):1219-1221 [FREE Full text] [doi: 10.1136/bmj.327.7425.1219] [Medline: 14630762]
51. McFall L. Personalizing solidarity? The role of self-tracking in health insurance pricing. Econ Soc 2019;48(1):52-76 [FREE Full text] [doi: 10.1080/03085147.2019.1570707] [Medline: 31057333]
52. Health intervention and technology assessment in support of universal health coverage. World Health Organization. 2014 May 24. URL: https://apps.who.int/iris/handle/10665/162870 [accessed 2020-07-03]

Abbreviations

GP: general practitioner
mHealth: mobile health

Edited by L Buis; submitted 27.08.20; peer-reviewed by A Lau, D Trupia, S Alomairah; comments to author 27.10.20; revised version received 15.03.21; accepted 17.05.21; published 09.08.21.

Please cite as:
Volpato L, del Río Carral M, Senn N, Santiago Delefosse M
General Practitioners' Perceptions of the Use of Wearable Electronic Health Monitoring Devices: Qualitative Analysis of Risks and Benefits
JMIR Mhealth Uhealth 2021;9(8):e23896
URL: https://mhealth.jmir.org/2021/8/e23896
doi:10.2196/23896
PMID:34383684

Original Paper

Assessing Electrocardiogram and Respiratory Signal Quality of a Wearable Device (SensEcho): Semisupervised Machine Learning-Based Validation Study

Haoran Xu[1], MSc; Wei Yan[2], MD; Ke Lan[3], MSc; Chenbin Ma[4], BSc; Di Wu[3], MSc; Anshuo Wu[5], BSc; Zhicheng Yang[6], PhD; Jiachen Wang[1], BSc; Yaning Zang[7], BM; Muyang Yan[2], PhD, MD; Zhengbo Zhang[8], PhD

[1]Medical School of Chinese PLA, Beijing, China

[2]Department of Hyperbaric Oxygen, The First Medical Center, Chinese PLA General Hospital, Beijing, China

[3]Beijing SensEcho Science & Technology Co., Ltd., Beijing, China

[4]School of Biological Science and Medical Engineering, Beihang University, Beijing, China

[5]University of Washington, Seattle, WA, United States

[6]PAII Inc., Palo Alto, CA, United States

[7]Department of Kinesiology, Shanghai University of Sport, Shanghai, China

[8]Centre for Artificial Intelligence in Medicine, Medical Innovation Research Department, Chinese PLA General Hospital, Beijing, China

Corresponding Author:
Zhengbo Zhang, PhD
Centre for Artificial Intelligence in Medicine
Medical Innovation Research Department
Chinese PLA General Hospital
28 Fuxing Road, Haidian District, Beijing
Beijing, 100853
China
Phone: 86 13693321644
Email: zhengbozhang301@gmail.com

Abstract

Background: With the development and promotion of wearable devices and their mobile health (mHealth) apps, physiological signals have become a research hotspot. However, noise is complex in signals obtained from daily lives, making it difficult to analyze the signals automatically and resulting in a high false alarm rate. At present, screening out the high-quality segments of the signals from huge-volume data with few labels remains a problem. Signal quality assessment (SQA) is essential and is able to advance the valuable information mining of signals.

Objective: The aims of this study were to design an SQA algorithm based on the unsupervised isolation forest model to classify the signal quality into 3 grades: good, acceptable, and unacceptable; validate the algorithm on labeled data sets; and apply the algorithm on real-world data to evaluate its efficacy.

Methods: Data used in this study were collected by a wearable device (SensEcho) from healthy individuals and patients. The observation windows for electrocardiogram (ECG) and respiratory signals were 10 and 30 seconds, respectively. In the experimental procedure, the unlabeled training set was used to train the models. The validation and test sets were labeled according to preset criteria and used to evaluate the classification performance quantitatively. The validation set consisted of 3460 and 2086 windows of ECG and respiratory signals, respectively, whereas the test set was made up of 4686 and 3341 windows of signals, respectively. The algorithm was also compared with self-organizing maps (SOMs) and 4 classic supervised models (logistic regression, random forest, support vector machine, and extreme gradient boosting). One case validation was illustrated to show the application effect. The algorithm was then applied to 1144 cases of ECG signals collected from patients and the detected arrhythmia false alarms were calculated.

Results: The quantitative results showed that the ECG SQA model achieved 94.97% and 95.58% accuracy on the validation and test sets, respectively, whereas the respiratory SQA model achieved 81.06% and 86.20% accuracy on the validation and test sets, respectively. The algorithm was superior to SOM and achieved moderate performance when compared with the supervised models. The example case showed that the algorithm was able to correctly classify the signal quality even when there were

complex pathological changes in the signals. The algorithm application results indicated that some specific types of arrhythmia false alarms such as tachycardia, atrial premature beat, and ventricular premature beat could be significantly reduced with the help of the algorithm.

Conclusions: This study verified the feasibility of applying the anomaly detection unsupervised model to SQA. The application scenarios include reducing the false alarm rate of the device and selecting signal segments that can be used for further research.

(JMIR Mhealth Uhealth 2021;9(8):e25415) doi:10.2196/25415

KEYWORDS

signal quality; electrocardiogram; respiratory signal; isolation forest; machine learning; mobile health

Introduction

Background

Wearable devices have been widely adopted for daily health care monitoring during the past decades. Many researchers utilize wearable sensors to continuously monitor physiological signals for mobile health (mHealth) and ubiquitous health (uHealth) app studies [1-3]. Recently, wearable devices have shown their potential in providing early warning of disease deterioration, chronic disease self-management, rehabilitation assessment, among others [4-7]. For example, some clinical deterioration changes in physiological signals could be often present 8-24 hours before a severe life-threatening event such as an unplanned intensive care unit admission or sudden cardiac death [8,9]. In these scenarios, signal quality is essential to acquire the valuable information from the time-series physiological signals which are very sensitive to noise. Signal quality assessment (SQA) facilitates reducing the high false alarm rate caused by signal quality [10] and can be applied to automatically screen the “real-world” data for further research. However, SQA of wearable physiological signals has not been well investigated. Such inadequate studies on signal quality reliability limit the further clinical deployment of these devices in the medical sector [11]. Therefore, it is important to develop a feasible method to evaluate the signal quality from wearable physiological monitoring systems and SQA is one of the basics of mHealth research and apps.

Related Work

It is widely recognized that the electrocardiogram (ECG) and respiratory signals are crucial for both patient monitoring and health status identification, and thus are being extensively investigated. Various solutions have been proposed to accomplish ECG SQA [12,13]. Some early studies, such as those by Langley et al [14] and Johannesen [15], considered the poor quality of ECG signals when their waveform features exceed the preset thresholds [16]. Several signal quality indices (SQIs) such as kSQI (the kurtosis of the distribution), sSQI (the skewness of the distribution), and pSQI (the relative power in the QRS complex) were introduced [17-19], which use the features from the time domain and the frequency domain of the ECG signals to assess the quality [20]. Another approach to ECG SQA is based on template matching. Researchers usually compare the similarity between the signals and a template that is fixed or derived from historical data [21]. In recent years, leveraging the machine learning technology in the medical domain, many researchers used the time–frequency domain features and SQIs to build machine learning models to achieve ECG SQA [16,21-23]. For example, Zhao et al [23] provided an algorithm based on convolutional neural networks, which aimed at identifying noisy segments from wearable ECG recordings. Zhang et al [16] compared the performance of random forest (RF), support vector machine (SVM), and their variants for ECG SQA with nonlinear features. For respiratory signals, Charlton et al [24] developed an SQI for the impedance pneumography respiratory signal by using the breath duration variations and by examining whether the peaks and troughs are clear and similarity of breath morphologies. However, research on respiratory SQA remains in its infancy. Few studies have investigated this topic so far to our knowledge.

Challenges

Owing to the rapid development of wearable devices, there is an explosion of the volume of data being acquired and available for research studies. However, the importance of the SQA process has been underestimated. The limitations of previous studies and the challenges we are currently facing are summarized as follows: For ECG SQA, first, signal quality is often judged subjectively, which lacks objective quantitative criteria, and the standard of signal quality was relatively fuzzy in previous studies [25,26]. Second, most of the SQAs were conducted under well-designed laboratory conditions by using simulated signals [27], or assessed the signals from bedside monitors. Thus, signals are highly different from those measured by wearable devices in daily lives because the noise in the laboratory was relatively single and controllable, or the signal quality was good for most of the time. Third, although most of the methods have good performances on ECG SQA, the dominant methods are still supervised machine learning models [16]. There is a concern that these models are at a high risk of overfitting, leading to unsatisfying model generalization. Moreover, when using supervised models, it is quite challenging to prepare tons of labeled data and even impossible for each research group to use the fixed open-source data sets, such as the MIT-BIH Arrhythmia Database (MITDB), to build models, which were not built for SQAs. In addition, hardware designs of wearable devices are diverse, resulting in aggravating incomplete generalization of data and poor migration performance of models. One possible solution to this problem is to build dedicated models using specific wearable devices and the data they collected. For respiratory SQAs, the challenge lies in the various respiratory patterns. Compared with ECG signals, respiratory signals have more diverse forms, broader spectral distribution, and different noise sources.

Study Objectives

To address the above problems, we pioneered the idea that the SQA process can be seen as an anomaly detection. The basic hypothesis of our study was that the decline of the signal quality can be quantified with the increase of the anomaly and can be detected by the machine learning model. The application scenarios we expected of the algorithm include reducing the false alarms caused by poor signal quality and selecting the high-quality signal segments for further research. The objectives and main components of this paper are to:

- design an algorithm based on the unsupervised machine learning model, isolation forest (IF), to classify the ECG and respiration signal quality into 3 different grades: good, acceptable, and unacceptable.
- quantitatively evaluate the performance of the algorithm on a small amount of labeled data. Further validation of the algorithm was implemented on several cases of data to prove its feasibility.
- apply the SQA algorithm to real-world data to demonstrate that the algorithm has the potential to reduce the false alarms caused by poor signal quality.

Methods

The Wearable Device and Data Sources

The medical-grade wearable device we used was a self-developed physiological signal monitoring system, SensEcho (Figure 1) [28], which has received clearance from the China Food and Drug Administration (CFDA) and has been deployed in the general wards of the Hyperbaric Oxygen (HBO) Department in Chinese PLA General Hospital (PLAGH) since 2018. The core wearable device of SensEcho is a vest, which provides a single-lead ECG signal, chest and abdominal respiratory signals via the respiratory inductive plethysmography (RIP) technology, and triaxial acceleration signals. It also allows for communication with other third-party wearable devices such as oximeters and blood pressure monitors. Its battery supports continuous monitoring for a minimum of 24 hours. For detailed information about SensEcho and the monitoring system, please refer to [29]. At the time of writing, SensEcho has collected more than 1000 records from patients and healthy individuals. Each record contains nearly 24-hour physiological signal monitoring results; thus, a large pool of data is available for research purposes. Data collection was carried out in a clinical environment for patients and from daily lives for healthy individuals without restriction of movement and activity. In this study, we used the single-lead ECG signal and chest respiration signal from the data pool to establish and evaluate the algorithm. This study was approved by the ethics committee of PLAGH (No. S2018-095-01).

Figure 1. Picture of SensEcho, including third-party oximeter and cuff blood pressure monitor.

Signal Quality Classification

Overview

The definition of signal quality was indistinct in previous studies, but some of the studies have proposed a few quantitative criteria. Inspired by [26] and the results of our pre-experiment, 10- and 30-second segments of ECG and respiratory signals were considered sufficient for our study. In early SQA studies, 5 quality groups (excellent, good, adequate, poor, and unacceptable) [15], 3 quality groups (acceptable, indeterminate, and unacceptable) [18,30,31], and 2 quality groups (acceptable and unacceptable) [32-35] were investigated. Based on previous studies, we defined 3 grades of signal quality for different requirements: (1) *good signal quality* refers to that in which the signal waves are clear, and signal of this grade can be analyzed automatically in follow-up studies and have confidence high enough for waveform feature analysis; (2) *acceptable signal quality* refers to that in which the R peak in ECG signal and peaks and troughs of respiratory signal can be accurately located by the algorithm, and the signal of this grade can be used for relative accurate heart rate and respiratory rate analysis. In addition, this grade is often the most difficult to distinguish and the signal availability depends on the specific apps where further manual determination might be needed; (3) *unacceptable signal quality* refers to that in which the waveform in the window is chaotic, and this grade of signal should be dropped because of the unreliable results obtained in signal analysis.

A brief description of characteristics of signal noise sources and their patterns is summarized in the following subsections [12,22,36,37].

Baseline Wander

ECG signals are affected by respiratory motion, body movement, and poor electrode contact. Respiratory signals are more sensitive to movement and breath pattern than ECG signals. One final major expression in signals is different levels of baseline wander.

High-Frequency Noise

For ECG signals, high-frequency noise usually includes power line interference, myoelectricity interference, and movement artifact. For respiratory signals measured by the RIP, the noise often is from vibrations caused by movement, such as moving or speaking.

Signal Loss

This is also a pervasive pattern in daily signal acquisition, which usually appears as a straight line. Based on the noise source and expression analysis, the quantitative evaluation criteria defined by clinical and engineering experts in our study are listed in Table 1.

Table 1. Quantitative signal quality assessment criteria.

Quality grade	Electrocardiogram	Respiratory signal
• Good	• ECG rhythm is clear; each QRS waveform can be distinguished with naked eyes. • No signal loss in the observation window. • Maximal baseline wander amplitude is less than one-third of signal amplitude in the observation window. • Pathological changes do not influence the signal quality assessment; the recognized obvious pathological patterns can be classified as good quality, such as ventricular premature beats.	• Regular waveform lasts for more than three-fourth of the observation window. • Maximal baseline wander amplitude is less than the signal amplitude in the observation window. • High-frequency noise can be easily filtered and does not affect the judgment of the respiratory signal waveform.
• Acceptable	• Low-intensity high-frequency noise; the R waves in signal can be recognized accurately. • No more than 2 high-frequency impulse noises occur in the observation window. • Less than 2-second signal loss in the observation window. • The maximal baseline wander amplitude is below the signal amplitude. • Fewer than 2 cardiac cycles in which the QRS waves cannot be recognized are allowed.	• One-half to one-fourth of the signal is clear; respiratory rhythm can be identified. • Time for signal loss or hold breath lasts less than one-half of the observation window. • High-frequency noise has only a little impact on the judgment of the overall waveform trend.
• Unacceptable	• Full of noise. • More than 2 R peaks in the observation window cannot be distinguished. • Excessive baseline wander. • Signal loss lasts more than 2 seconds. • Suspected pathological changes, but the cause is not clear.	• The pattern of respiratory waveform is difficult to recognize. • Severe baseline wander.

Isolation Forest

IF is an unsupervised anomaly detection model that has been applied to many fields such as streaming data processing and mineral mapping [38,39]. IF grows an ensemble of binary trees to estimate the degree of being an anomaly of an instance. As anomalies are more susceptible to isolation, they have a short path length [38,40]. Furthermore, an anomaly score can be

obtained by measuring/estimating the average path height of the ensemble of binary trees (in [40], the authors named them iTree). The IF model is based on 2 fundamental assumptions and premises. The first one is that the anomalies should be "few and different." If a pattern occurs frequently in the training set, it will be more likely to be perceived as normality, although it is indeed an anomaly manually determined. The second one is that the training set should conclude as many normal patterns of the signals as possible. It is necessary to guarantee that the training set has a large enough variety, especially for normal signals; otherwise the model will be more likely to classify a brand-new pattern as an anomaly.

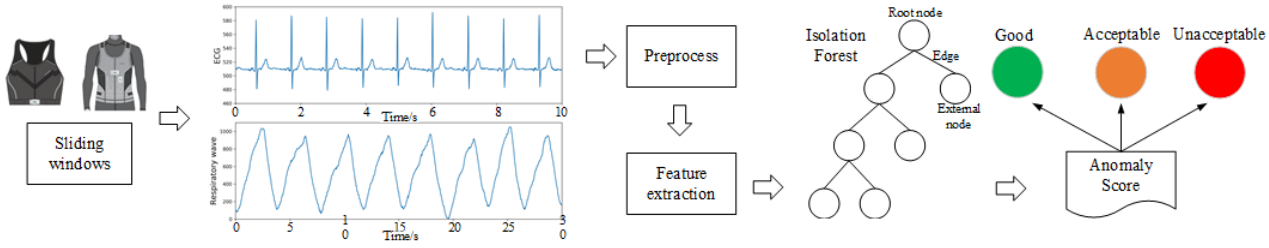
Based on the above theory, the general framework of the SQA algorithm is shown in Figure 2. We built models for ECG and respiratory SQA, respectively, and both models were trained and evaluated independently. The preprocess included filtering, removing the outliers, removing the baseline, and normalization. We then selected 8 and 18 features from the time and frequency domains of the ECG and respiratory signals, respectively. Skewness, kurtosis, and distances of adjacent waveforms calculated using the dynamic time warping method [41] were the key features we used, which also have been widely adopted as the key variables to construct the SQIs [17,18,42,43]. The skewness and kurtosis are defined as Equations (1) and (2). Other features we used in this study were the features from amplitude of the signal in the time–frequency domain, power spectrum distribution, and power spectral density.

(1)

(2)

where N is the sample points of the signal, is the mean value, and σ is the SD.

Figure 2. General framework of the signal quality assessment algorithm for the electrocardiogram and respiratory signal.



Experiment Design

Overview

The experimental process involved 4 key steps. The model training and validation were conducted on 4 nonoverlapping data sets extracted from the sizable volume data pool and possessed different functions: (1) training set, which was used to train the IF model; (2) validation set, which was used to find the thresholds that map the anomaly scores obtained by the model to the triclassification SQA results; (3) test set, which was used to quantitatively measure the generalization ability of the model; and (4) case set, which was used to qualitatively evaluate the model's performance by feeding a whole case of data to it. Some details of these 4 data sets are specified in the following sections.

Training Set

We selected a set of 24-hour monitoring records which met the following inclusion criteria: (1) signal acquisition was stable by manual determination; (2) no signal loss for extended periods (over 10 minutes) during monitoring; and (3) no persistent atrial fibrillation during monitoring. Based on these, 30 records were included and we selected 3-10 of them randomly to construct the training set with their whole data. We repeated the selection process 20 times for each epoch, that is, we randomly selected 3 records to construct the training set 20 times to find the best performance of the model.

Validation Set

We used the data from 16 patients and 8 healthy individuals to construct this data set, expecting that the pathological changes were more complex and the proportion of anomaly was relatively high. We selected 10,000 windows of signals from the records and then removed half of them that were obviously of high quality. The data set was labeled independently by 3 pretrained graduate students of biomedical engineering according to the criteria in the above section. To guarantee label accuracy, we used the agreed result to define the final label, and dropped the windows of signals that had conflicting label results. Moreover, we asked clinical specialists to mark whether the ECG signals in the data set were pathological. If pathological manifestations of the signal, such as arrhythmia or ST-segment elevation, were confirmed, the number of this signal segment was recorded additionally. After the manual annotation of the data set is completed, the anomaly scores of the labeled data can be obtained by feeding the signals to the trained SQA model. Then, thresholds T1 and T2 were set to map the anomaly scores to the signal quality grades. We adjusted the values of T1 and T2, respectively, to find the best performance thresholds, which were fixed and used in the next step.

Test Set

Test set data came from 8 patients and 9 healthy individuals, because we expected the test set to be somewhat different from the validation set and to be closer to practical use. We extracted 1 window of signals every 6 minutes and this data set initially comprised 5500 windows of signals, which were labeled in the same way as the validation set. We used the T1 and T2 values

determined by the validation set to obtain the classification results of the model, and then quantitatively evaluated the generalization ability of the model. The basic information about the individuals involved in the validation and test sets is summarized in Table 2.

Table 2. Basic information about the individuals utilized in the validation and test sets.

Characteristics	Validation set		Test set	
	Patients (n=16)	Healthy individuals (n=8)	Patients (n=8)	Healthy individuals (n=9)
Demography				
Male, n (%)	9 (56)	8 (100)	5 (63)	5 (56)
Age (year), mean (Q1-Q3)	56 (52-60)	27 (25-33)	69 (65-73)	32 (27-41)
Height (cm), mean (Q1-Q3)	168 (160-170)	174 (171-176)	165 (156-174)	171 (157-175)
Weight (kg), mean (Q1-Q3)	68 (55-76)	68 (59-74)	70 (64-78)	73 (58-74)
Comorbidity, n (%)				
Coronary heart disease	12 (75)	—	5 (63)	—
Hyperlipemia	9 (56)	—	4 (50)	—
Hypertension	9 (56)	—	7 (88)	—
Diabetes	8 (50)	—	4 (50)	—
Pulmonary nodule	4 (25)	—	2 (25)	—
Sleep apnea syndrome	2 (13)	—	—	—

Case Set

We fed several cases of data to the model. Different grades of signal quality segments were marked in different colors. We looked at several observation windows in detail to determine whether the model classification results were correct. Note that we are particularly concerned about the pathological changes in the cases, because we expected pure pathological changes to be not misclassified as poor signal quality.

Data Set Descriptions

After data labeling, we obtained the final validation and test sets. The validation set consisted of 3460 and 2086 ECG and respiratory labels (all agreed), respectively. Of the 3460 ECG labels, 3022 (87.34%) were good, 189 (5.46%) were acceptable, and 249 (7.20%) unacceptable. Of the 2086 respiratory labels, 1308 (62.70%) were good, 511 (24.50%) acceptable, and 267 (12.80%) unacceptable. The test set consisted of 4686 and 3341 ECG and respiratory labels, respectively. Of the 4686 ECG labels, 3767 (80.39%) were good, 284 (6.06%) acceptable, and 635 (13.55%) unacceptable, compared with 2255 (67.49%), 587 (17.57%), and 499 (14.94%), respectively, for respiratory labels. Some typical examples of the labeled ECG and respiratory signals are shown in Figures 3 and 4.

Meanwhile, for the pathological ECG labels, a total of 661/3460 (19.10%) windows of ECG signal in the validation set were marked. Of these, 648 (98.0%) were labeled as having good quality and the rest (13/661, 1.9%) as acceptable quality. In the test set, 634/4686 (13.53%) windows of signal were pathological; of these, 618 (97.5%) were of good quality and the rest (16/634, 2.5%) were of acceptable quality.

Figure 3. Typical examples of the labeled electrocardiogram signals. (a) & (b) are the normal, good-quality signals; (c) is suspected of arrhythmia while (d) is an expression of ventricular premature beats (VPBs); (e) – (h) show examples of baseline wander, power line interference and impulse noise; (i) – (k) show examples of severe noise and signal loss; (l) is suspected of VPBs but the signal is unclear.

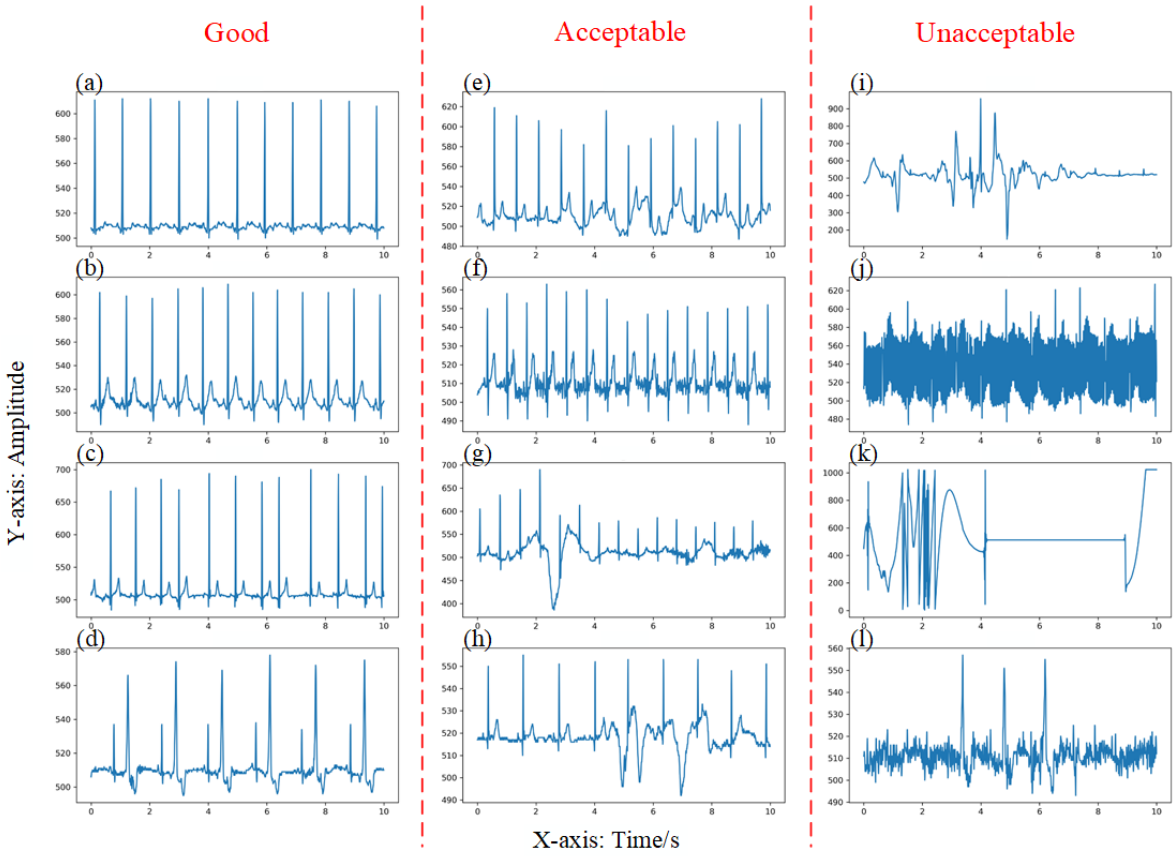
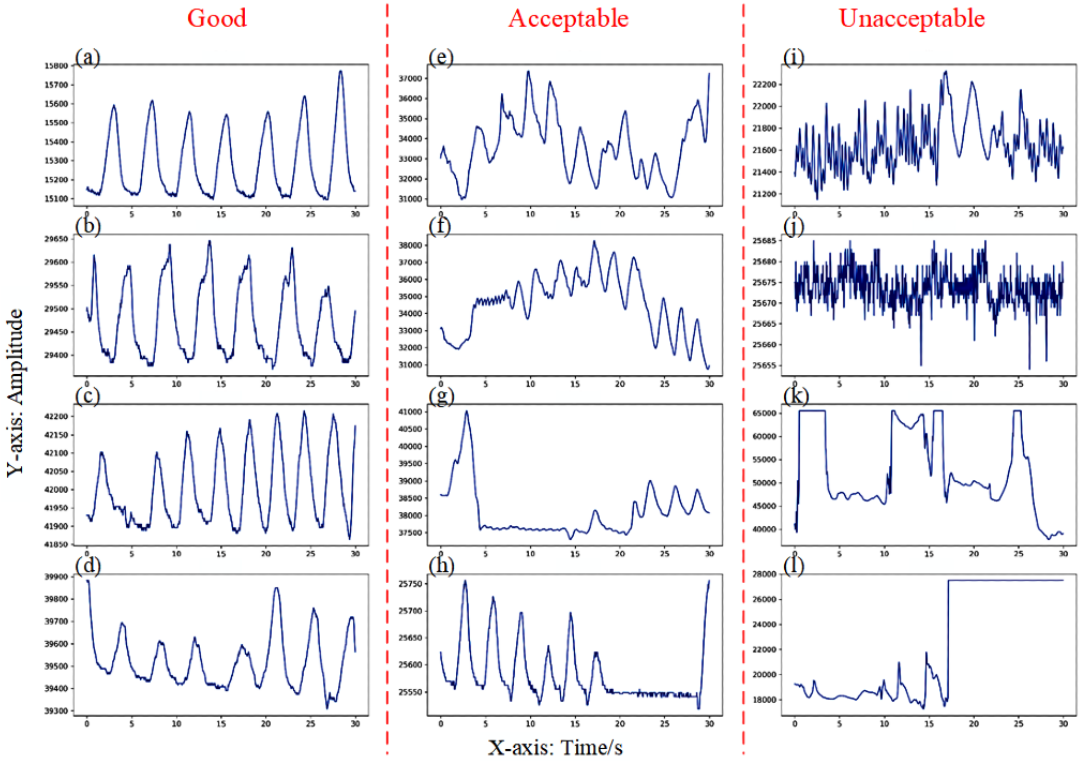


Figure 4. Typical examples of the labeled respiratory signals. (a) – (d) show clear and regular respiratory waves; signals in (e) – (h) do not have enough regularity, apnea occupies some small segments in the observation windows; (i) – (l) show severe noise and signal loss in the observation windows.



Performance Evaluation

The programming language we used was Python (version 3.6.5) and the major library in this study is scikit-learn (version 0.23.1). The proposed algorithm contained 2000 trees and had 5% anomaly proportion as parameters. We first evaluated the algorithm's performance according to its accuracy score, which is defined as the number of correctly classified samples divided by the total number of samples. Some additional evaluation indicators included mean precision rate, recall rate, and F1 score (marco-F1). To further evaluate the performance of the algorithm, we compared the algorithm with the self-organizing maps (SOMs) [44] and 4 classical supervised machine learning models, namely, logistic regression (LR), SVM, RF, and extreme gradient boosting (XGB). It should be noted that the SOM is an unsupervised model based on artificial neural network and has been applied in several health care–related signal processing fields such as photoplethysmogram signal classification [45,46] and health situation monitoring [47,48]. The SOM library used in this study was MiniSom (version 2.2.7) and the SOM model was trained using 10,000 interactions and a 10 × 10 grid on the training set with the learning rate of 0.05. For RF, we used 1000 trees, whereas for XGB, we chose the following hyperparameters: "binary: softmax" as the logistic function and "approx" as the tree method. The other parameters of the models were default. Features were normalized before being fed to LR, SVM, and SOM.

According to our evaluation strategy, for unsupervised models, we trained the models on the training set and found best thresholds on the validation set. For supervised models, we trained the models on the whole validation set. We then compared the performance of both supervised and unsupervised models on the test set. The accuracy, precision, recall, and F1 scores are calculated.

We also investigated the performance of the proposed model with fewer labels in comparison with that of the reference model. We randomly selected 200, 600, and 1000 labels in the validation set to find the thresholds for the unsupervised models and train the supervised models, and then test these on the whole test set. Each random selection is repeated 30 times, and then the mean and SD of the accuracy of the models are computed.

Algorithm Application

We applied the designed SQA algorithm to 1144 cases of data collected in the HBO Department of PLAGH; each of the cases had a dynamic ECG record of nearly 24 hours. Each record of data was read by a clinical expert to give an overall signal quality evaluation result. According to the results, the data were divided into 3 groups, representing different grades of quality of the whole signals. We also scanned the data with an arrhythmia detection algorithm, which is commonly used in automatic dynamic ECG analysis, and the real-time alarm function of SensEcho. The core technology of the arrhythmia detection algorithm is traditional signal processing methods, including filtering and wavelet decomposition. We learned about the type, onset, and duration of each arrhythmia alarm detected by the arrhythmia detection algorithm. For the purpose of this study, a false alarm was defined as the onset of 1 arrhythmia alarm marked with poor signal quality. The proportion of different quality of signals, the number of various arrhythmia alarms, and the percentage of false alarms in each group were calculated.

Results

Model Performance

For the training set that is important for the IF model, we randomly selected monitoring records as described in the "Experiment Design" section and built the training sets to train the model to guarantee the variety and find the best performance of the model. Quantitative evaluation results of the model performance on the validation and test sets are shown in Figure 5. For ECG signals, the model performed at the same level on both validation and test sets, but for respiratory signals, the model performed slightly better on the test set than on the validation set. This is reasonable because the two data sets were constructed differently; thus, the test set was easier for SQA classification. Models that performed the best on the test set were selected for further study. The scores gained from the best model for ECG SQA and the best classification thresholds are shown in Figure 6, in which the accuracy reached 94.97% and 95.58% on the validation and test sets, respectively. The confusion matrixes are shown in Figure 7. Similarly, the scores for respiratory SQA and the thresholds are shown in Figure 8. This model achieved 81.06% and 86.20% accuracy on the validation and test sets, respectively. Figure 9 shows the confusion matrix of the results.

The results regarding the classification efficiency of the pathological ECG signal are summarized as follows: in the validation set, 100% (648/648) of good-grade and 23% (3/13) of acceptable-grade pathological ECG signals were classified correctly; however, 77% (10/13) of acceptable-grade signals were misclassified as good quality. In the test set, 99.8% (617/618) of good-grade and 31% (5/16) of acceptable-grade pathological signals were classified correctly; however, 1 sample of good-quality signal was misclassified as acceptable grade and 69% (11/16) of acceptable-grade signals were misclassified as good quality. The above results showed that the model also had a good classification effect on pathological signals: In this study, the vast majority of pathological signals were correctly classified and the misclassification will not increase false-negative decisions.

Figure 5. Quantitative evaluation of the model performance on the validation set and test set. ECG: electrocardiogram.

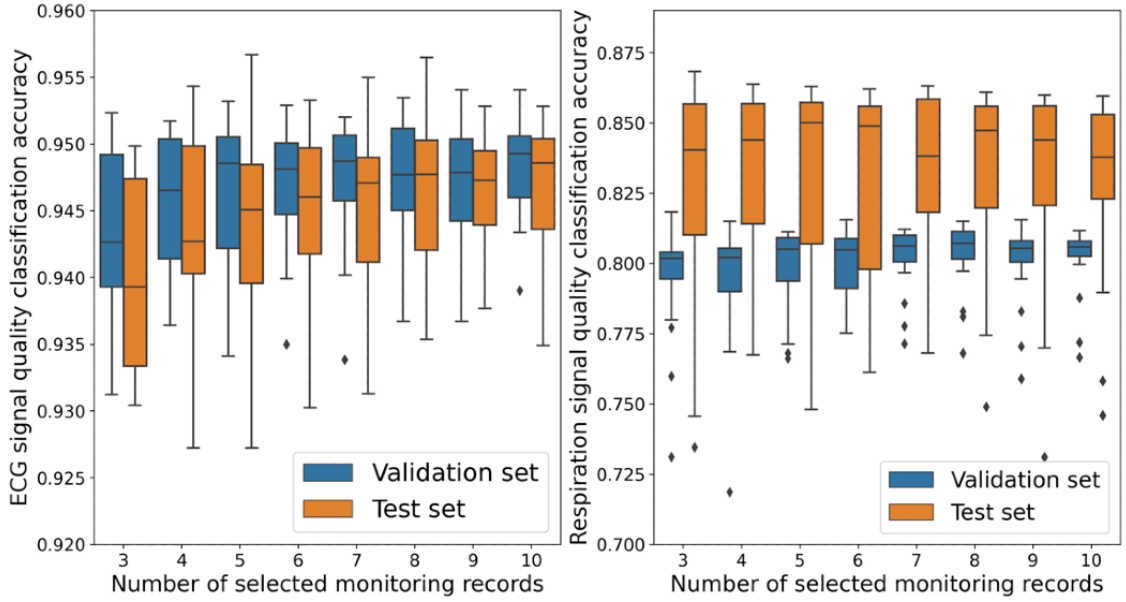


Figure 6. Electrocardiogram (ECG) signal anomaly scores on the validation set and test set, and the best performance thresholds.

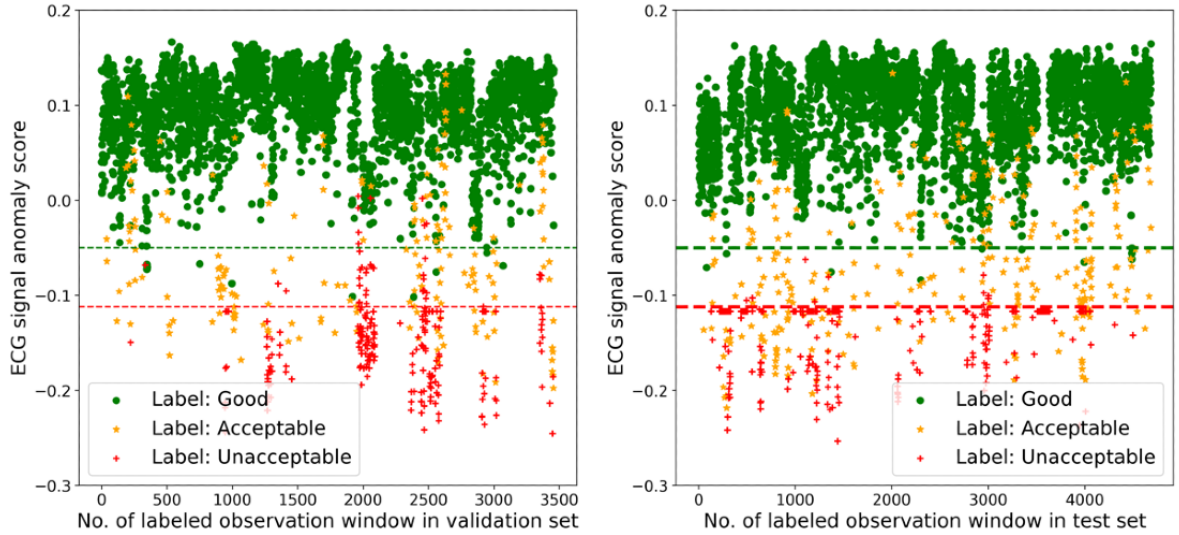


Figure 7. The electrocardiogram confusion matrixes of the results. 0: Good; 1: Acceptable; 2: Unacceptable.

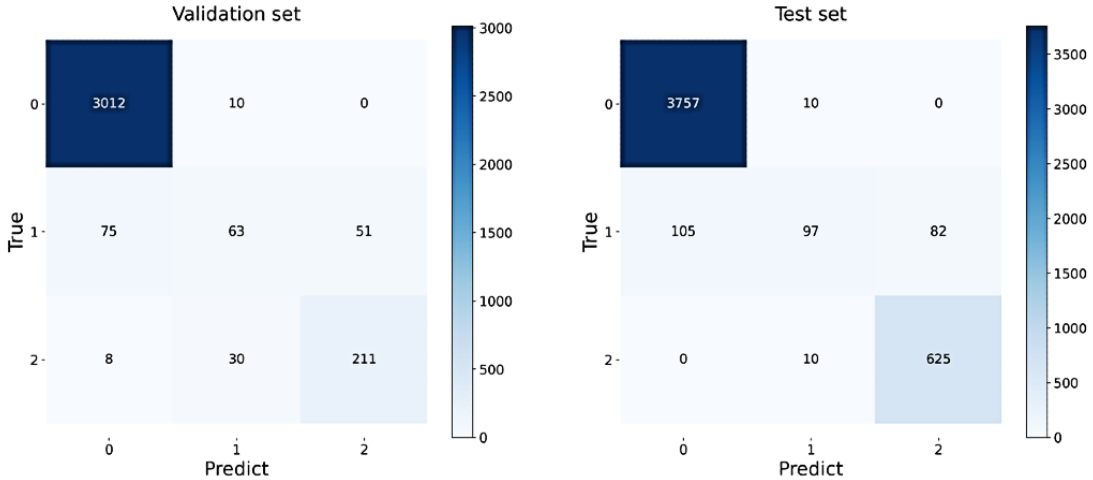


Figure 8. Respiratory signal anomaly scores on the validation set and test set, and the best performance thresholds.

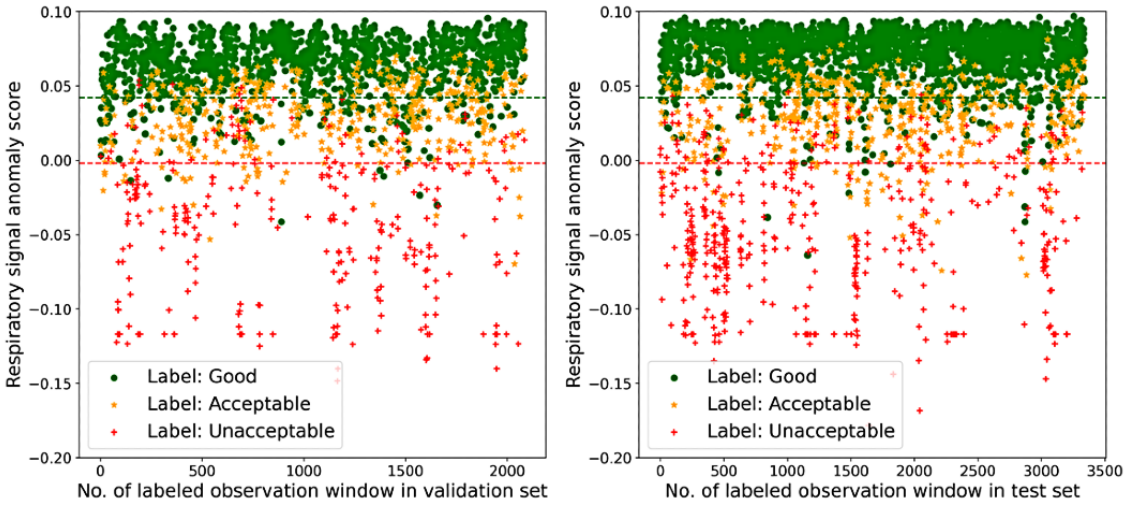
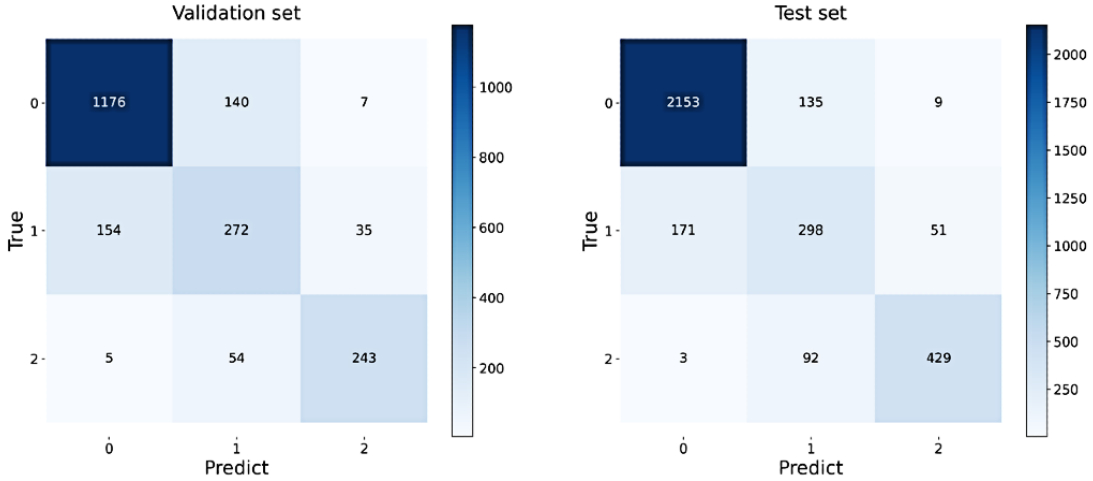


Figure 9. The respiratory confusion matrixes of the results. 0: Good; 1: Acceptable; 2: Unacceptable.



Performance Evaluation Results

The classification results of the desired algorithm and reference models of the test set are summarized in Tables 3 and 4. From Table 3, it can be found that, for supervised models, the LR model performed the worst for both ECG and respiratory signals. Meanwhile, RF and XGB performed slightly better than the proposed algorithm. Understandably, supervised models generally have better performance than unsupervised models. For unsupervised models, SOM performed worse than the proposed model. For ECG SQA, the SOM achieved 0.91 accuracy and 0.55 F1 score on the validation set, indicating an insufficient generalization ability of the thresholds in this scenario for the model. We speculated that the complex pathological changes and noise in the data set made it difficult for SOM to perform dimensionality reduction and correctly map the model outputs to the SQA results. From Table 4, it can be found that the proposed model had a better performance when the number of labels is small. When the number of labels is greater than 1000, the performance of the supervised models was better than that of the proposed model. In other words, when we do not have enough labeled data, the unsupervised model is superior. However, we still recommend preparing slightly more labels as possible to guarantee the stability and generalization ability of the thresholds.

Table 3. Model performance on the test set.

Model	Electrocardiogram				Respiratory signal			
	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
Supervised models								
Logistic regression	0.79	0.79	0.61	0.50	0.79	0.72	0.55	0.59
Support vector machine	0.96	0.93	0.77	0.79	0.80	0.72	0.57	0.60
Random forest	0.97	0.95	0.84	0.87	0.92	0.88	0.85	0.87
Extreme gradient boosting	0.97	0.95	0.86	0.89	0.91	0.86	0.85	0.85
Unsupervised models								
Self-organizing maps	0.82	0.57	0.39	0.40	0.77	0.65	0.51	0.51
Isolation forest, proposed unsupervised model	0.96	0.90	0.77	0.80	0.86	0.79	0.78	0.78

Table 4. The accuracy on the test set of models with fewer labeled data.

Number of labels	Logistic regression	Support vector machine	Random forest	Extreme gradient boosting	Self-organizing maps	Isolation forest, proposed unsupervised model
ECG[a], mean (SD)						
200	0.80 (0.00)	0.85 (0.05)	0.86 (0.06)	0.84 (0.06)	0.80 (0.01)	0.89 (0.06)
600	0.80 (0.00)	0.86 (0.05)	0.89 (0.06)	0.88 (0.05)	0.81 (0.01)	0.90 (0.06)
1000	0.81 (0.00)	0.90 (0.04)	0.93 (0.04)	0.92 (0.04)	0.81 (0.01)	0.93 (0.02)
Respiratory signal, mean (SD)						
200	0.71 (0.05)	0.71 (0.06)	0.80 (0.04)	0.79 (0.03)	0.70 (0.02)	0.82 (0.04)
600	0.75 (0.04)	0.75 (0.06)	0.85 (0.02)	0.84 (0.02)	0.73 (0.03)	0.84 (0.02)
1000	0.77 (0.04)	0.76 (0.06)	0.87 (0.01)	0.86 (0.01)	0.72 (0.07)	0.85 (0.01)

[a]ECG: electrocardiogram.

Case Validation

To further evaluate the performance of the algorithm on SQA, the algorithm was tested on several cases. In this paper, ECG and respiratory signals of a patient are illustrated. The patient is a 65-year-old male, standing 170 cm tall, and weighing 68 kg when admitted, and had been monitored by the SensEcho in the general ward of the HBO Department. He was diagnosed with coronary heart disease, posterior mitral valve prolapse, hypertension risk level 2, hyperuricemia, and fatty liver disease.

As shown in Figures 10 and 11, the different signal quality grades classified by the algorithm were marked in 3 colors: the green segments stand for the good quality, the yellow segments for the acceptable quality, and the red segments for the unacceptable quality. Furthermore, in these figures, 4 windows of the monitoring signals were selected to elaborate and illustrate the detailed signals and the classification results, respectively. It can be seen that the monitoring lasted for up to 24 hours, but there was not much high-quality data available in this case. Signal loss was the most common unacceptable signal quality expression and the segments were all marked in red. ECG and respiratory signals of the last few hours were full of noise, so it was suspected that the patient might have removed the device ahead of time.

We found that the pathological changes in ECG did not influence the SQA process directly (Figure 10). Most of the observation windows with ventricular premature beats (VPBs) were also marked in green and yellow correctly, that is, in this case the pathological changes were not filtered which met our expectations. In Figure 11, acceptable and unacceptable signal quality segments are more numerous and dispersed for respiratory signals compared with ECG signals. The good-quality segments were mainly concentrated during the patient's bed rest period, as breath was more controllable and vulnerable to noise during the day. In conclusion, the algorithm demonstrated an excellent performance in this case and it can be used to automatically screen out the good-quality segments for further research.

Figure 10. A signal quality assessment case example of the whole monitoring 24-hour electrocardiogram signal (Green: Good segments; Yellow: Acceptable segments; Red: Unacceptable segments).

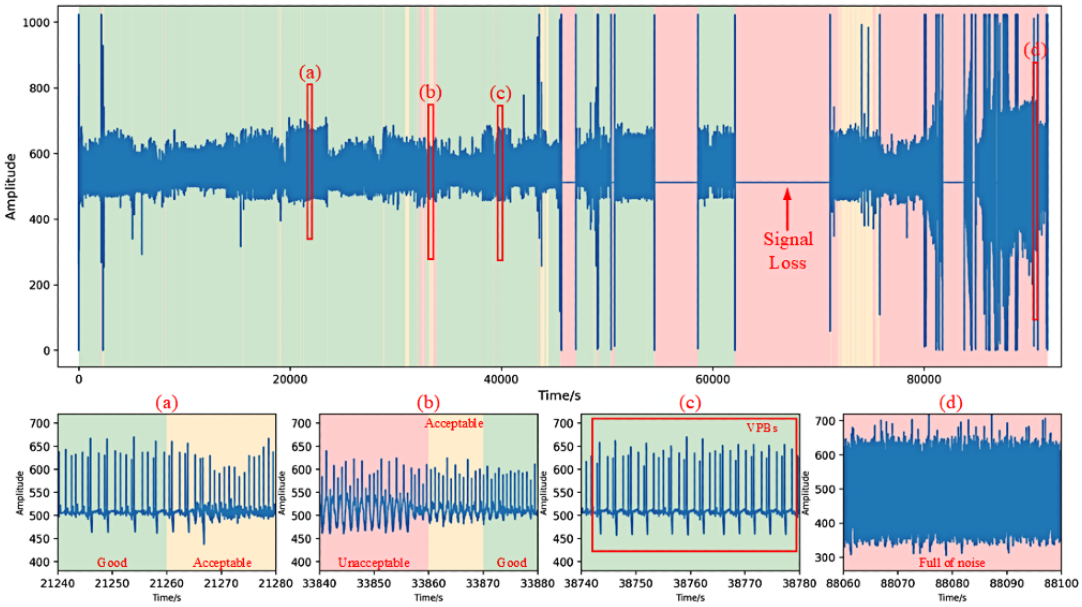
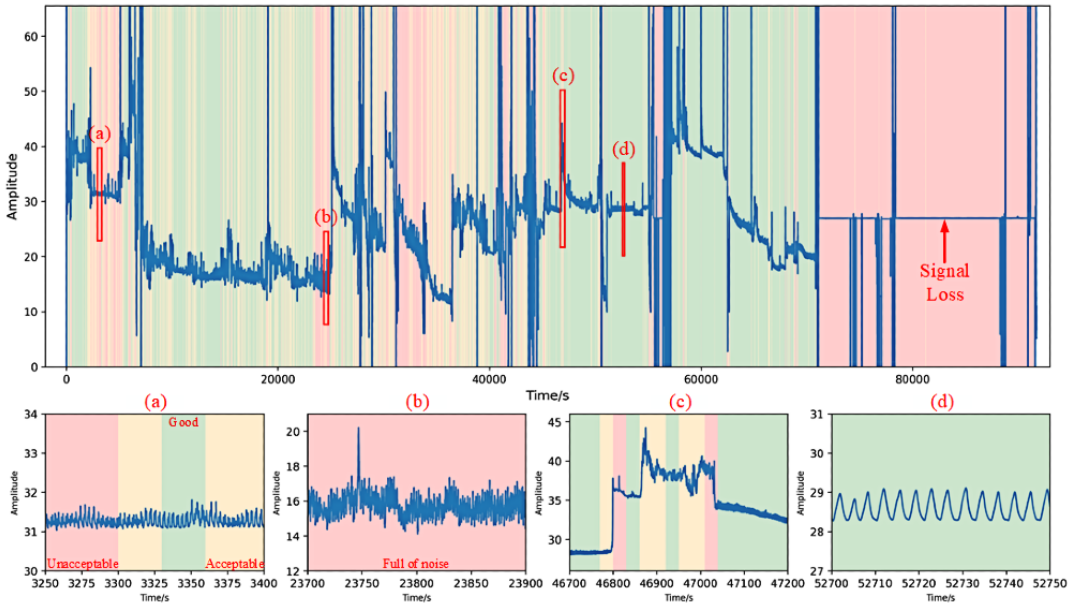


Figure 11. A signal quality assessment case example of the whole monitoring 24-hour respiratory signal (Green: Good segments; Yellow: Acceptable segments; Red: Unacceptable segments).



Algorithm Application Results

The algorithm application results are summarized in Table 5. The types of arrhythmia alarm we were concerned about were bradycardia, tachycardia, atrial premature beat (APB), VPB, atrial bigeminy, and atrial trigeminy. The "count" column represents the number of cases with a specific arrhythmia alarm detected; for example, bradycardia was detected in 525 cases out of the total 1144 cases. From Table 5, it can be seen that the age, weight, and height of the 3 groups of patients were basically on the same level, whereas the proportion of females increased in the medium and worst groups, indicating that the quality of ECG signal measured from female users might be poor due to hardware. The proportion of different signal quality grades in these cases means that the best group of patients has the highest percentage of good quality and the lowest percentage of unacceptable quality, whereas the worst group of patients has the lowest percentage of good quality and the highest percentage of unacceptable quality. Among these cases, the median [Q1-Q3] for good, acceptable, and unacceptable quality proportion was 90.0% [81.4%-95.9%], 4.8% [2.1%-8.0%], 4.0% [1.1%-9.3%], respectively. These results have 2 implications: First, the desired SQA algorithm is consistent with the common knowledge of people, which can be used to analyze the quality of signals measured by SensEcho automatically and quantitatively. Second, the vast majority of ECG signals measured by SensEcho are usable, which demonstrates that the

wearable device can effectively monitor patients' ECG signal for most of the time.

For the arrhythmia alarm results, ideally, the number of various arrhythmia alarms within each group should be similar. However, it was observed that the number of APBs and VPBs increased significantly (P=.02 and <.001, respectively), suggesting that the signal quality did affect the accuracy of the arrhythmia detection algorithm and that some of the alarms might have been caused by poor signal quality. For the defined false alarm results, the APBs and VPBs increased significantly (P<.001 for both) in the medium and worst groups, and the false alarm of VPBs even accounted for 60.4% [23.9%-87.3%] in the worst group, compared with 18.2% [0.0%-61.5%] for the VPBs among all cases. In addition, it was found that tachycardia had a very high false alarm proportion, probably due to the movement of patients with poor signal quality. We considered that the aforementioned types of false alarms can be detected and effectively reduced by the desired SQA algorithm. Meanwhile, it was also found that for some types of arrhythmia alarms such as those for atrial bigeminy and atrial trigeminy, the arrhythmia detection algorithm was accurate and rarely affected by the signal quality.

Table 5. Results of the SQA algorithm and the arrhythmia detection algorithm applied to the data collected from the Hyperbaric Oxygen Department.

Characteristic	Grouped by manual evaluation			Total (n=1144)	Count
	Best (n=671)	Medium (n=365)	Worst (n=108)		
Demography					
Female, n (%)	239 (35.6)	145 (39.7)	45 (41.7)	429 (37.5)	1144
Age (year), median (Q1-Q3)	59.3 (53.0-66.8)	61.5 (54.0-67.9)	60.6 (53.5-66.7)	60.1 (53.5-67.1)	1144
Weight (kg), median (Q1-Q3)	70.0 (62.0-78.5)	71.0 (65.0-80.0)	70.0 (63.0-81.0)	70.3 (63.0-80.0)	1144
Height (cm), median (Q1-Q3)	168.0 (160.0-173.0)	168.0 (160.0-173.0)	167.5 (160.0-174.0)	168.0 (160.0-173.0)	1142
Proportion of different signal quality grades detected by the algorithm (%), median (Q1-Q3)					
Good	93.2 (87.0-97.3)	85.9 (78.6-92.0)	75.5 (61.9-86.2)	90.0 (81.4-95.9)	1141
Acceptable	3.8 (1.7-7.0)	5.8 (3.2-9.4)	5.8 (4.3-9.0)	4.8 (2.1-8.0)	1141
Unacceptable	2.1 (0.6-5.4)	7.0 (2.9-12.7)	15.2 (6.6-29.2)	4.0 (1.1-9.3)	1141
Arrhythmia alarm count, median (Q1-Q3)					
Bradycardia	4.0 (2.0-7.0)	3.0 (2.0-6.0)	2.0 (1.0-4.5)	4.0 (2.0-6.0)	525
Tachycardia	1.0 (1.0-2.0)	1.0 (1.0-2.0)	1.0 (1.0-2.0)	1.0 (1.0-2.0)	224
APB[a]	11.0 (4.0-34.5)	15.0 (6.0-42.0)	17.0 (4.0-46.0)	13.0 (5.0-39.0)	1103
VPB[b]	6.0 (2.0-25.2)	14.0 (4.0-54.0)	25.0 (5.0-76.0)	9.0 (3.0-39.0)	987
Atrial bigeminy	4.0 (1.0-13.0)	5.0 (2.0-9.0)	2.5 (2.0-6.0)	4.0 (2.0-10.0)	79
Atrial trigeminy	4.0 (2.0-10.5)	5.0 (1.0-10.8)	6.0 (2.2-11.2)	4.5 (1.8-11.0)	88
Defined false alarm proportion (%), median (Q1-Q3)					
Bradycardia	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	525
Tachycardia	50.0 (0.0-100.0)	100.0 (0.0-100.0)	100.0 (50.0-100.0)	100.0 (0.0-100.0)	224
APB	0.0 (0.0-9.9)	5.6 (0.0-28.6)	14.1 (0.0-70.6)	0.4 (0.0-19.1)	1103
VPB	6.2 (0.0-50.0)	35.6 (1.9-71.7)	60.4 (23.9-87.3)	18.2 (0.0-61.5)	987
Atrial bigeminy	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	79
Atrial trigeminy	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	88

[a]APB: atrial premature beat.

[b]VPB: ventricular premature beat.

Discussion

Contributions and Principal Findings

Our highlights and key contributions are summarized as follows:

- We achieve the ECG and respiratory SQA by using an unsupervised model, IF, which has not been applied in SQA before. Furthermore, we attempted to verify the idea that the SQA process can be viewed as an anomaly detection. In this study, the proposed algorithm was superior than SOM and achieved moderate performance when compared with the supervised models.
- We applied the SQA algorithm to a large data set with 1144 records of ECG signal. The results demonstrate that the arrhythmia alarm accuracy could be influenced by the signal quality, and the SQA algorithm has the potential to reduce

some specific types of arrhythmia false alarms such as tachycardia, APB, and VBP caused by poor signal quality.
- To our knowledge, this is one of the earliest studies that focuses on the quality of respiratory signals measured via the RIP technology. It provides a method to automatically select the high-quality segments of respiratory signal for further studies.

One featured point in our study is that 3 data sets that have different functions were used to construct and quantitatively validate the algorithm. In the workflow of our study, the training set was a large volume data set in which ideally all the patterns of the signal could be enumerated, while the validation set and the test set were unseen by the model when we trained it. We also conducted a very small experiment, where we directly trained the models on the validation set, found the best performance thresholds, and then evaluated the performance of the models on the test set. The results showed that for the ECG signal, the model achieved 0.92 accuracy and 0.72 F1 score, whereas for the respiratory signal, the model achieved 0.72 accuracy and 0.68 F1 score, which are lower than the current performance in the "Results" section. These results demonstrate that the diversity of patterns in the training set ensures the generalization performance of the unsupervised model. In fact, in an era of big data, it is easy to obtain a training set with a large sample size, yet lacking labels. The workflow we proposed in this study provides a feasible way to take advantage of the large sample size that can be applied in follow-up studies.

What should be emphasized is that we included the respiratory signal measured via RIP in this study for 2 reasons. First, the respiratory signal is an important physiological signal, which contains abundant personalized information, indicating the health status and disease deterioration of a person. More importantly, the quality of respiratory signal measured via RIP is not well investigated compared with ECG. In our study, we would like to point out that a signal with relatively little research and no fixed waveform could also be assessed by this method, which has the potential to be extended to other SQA scenarios such as impedance pneumography respiratory signal, dynamic blood pressure, and photoplethysmogram. That is, our study provides a practical workflow for other time-series physiological signal research groups to develop their own applicative SQA algorithms.

Limitations

There are also some limitations to our work. First, the model we used was an unsupervised machine learning model, which lacks enough interpretability and the performance is largely determined by the quality of the training set. We attempted several construction methods of the training set, yet it was hard to guarantee that the models achieved the best performance. Second, the classification results of the models for the medium grade of signal quality were not good. The sensitivities of the algorithm for this grade are only 0.34 for ECG and 0.57 for respiratory signals, respectively, which seriously lower the overall F1 scores of the models. This is because the medium level of signal quality is always the hardest to classify even manually. We tried some approaches such as data augmentation and constructing an artificial training set. However, the results showed no significant improvement. It is worth mentioning that the SOM showed moderate performance in the unsupervised methods, perhaps because, in our study, the framework, especially the training and generalization methods, was not suitable for this model. Further to this point, SOM and the rapidly evolving deep learning methods are worth being investigated after further accumulation of data. Third, as the validation of the algorithm on pathological signals was insufficient, although the results in this study were good, we still consider that the algorithm has the risk of misclassifying pathological changes as abnormal as a result of noise. We thus need to further validate the algorithm, which demands more pathological data accumulation and long-term feedback of actual use from clinicians.

Future Work

Our future research includes the following. First, the algorithm calls for more comprehensive experimental validation. Accordingly, we should further verify the performance of the model in the presence of pathological changes and quantify how much the model can reduce the false alarm rate. It requires long-term usage and more data collection, especially from patients with specific diseases such as arrhythmia and chronic obstructive pulmonary disease. Second, we will test the time usage and real-time performance of the algorithm. To our knowledge, the IF model operation does not take too much time when the thresholds are determined, yet the feature extraction process is more time-consuming. As we preliminarily tested, the whole SQA process for ECG signal takes 0.3-0.5 seconds on server for every observation window (10 seconds). For respiratory signal, it takes less than 0.1 seconds for every observation window (30 seconds). We will integrate the algorithm into the server to achieve the real-time SQA. Third, there are many mHealth and uHealth apps nowadays, but there is a lack of assessment of the data measured under nonlaboratory conditions and their usability. Based on the algorithm we developed, we will further evaluate the value of the wearable device, SensEcho, in daily life situations from a signal quality perspective, find the cause of the decrease in signal quality, and improve the both hardware and software of the wearable device. We believe that this will further promote the application of mHealth and uHealth.

Conclusions

In this study, the results verified our hypothesis that the SQA problem can be seen as an anomaly detection. We built a model based on the unsupervised machine learning model, IF, to avoid heavy data annotation work and to realize ECG and respiratory SQA. What distinguishes us from other studies that used the IF model is that we used a small amount of labeled data to enable the mapping of model scores to human cognitive classification results. Our validation results indicate that the proposed algorithm is superior than SOM and shows a moderate performance compared with supervised models. Meanwhile, the proposed algorithm has the advantages of flexibility, easy adjustment, and better performance with few labeled data. In addition, the pathological changes in our case are correctly classified, demonstrating the model's good application effect. The algorithm application results on 1144 cases from the clinic

suggest that the proposed algorithm has the potential to reduce some types of arrhythmia false alarms such as tachycardia, APB, and VBP.

Middle-aged and elderly people, such as patients in the HBO Department in this study, often suffer from complex chronic diseases and are at relatively high risk even in hospitals. Therefore, the adoption of wearable devices in clinics and the advancement of data analysis could provide easily accessible health care that can greatly benefit this population. We consider that the proposed algorithm can advance the clinical apps of wearable devices and facilitate follow-up mHealth and uHealth studies of various time-series physiological signals.

Acknowledgments

This work was supported by NSF of China (61471398), Beijing Municipal Science and Technology (Z181100001918023), Special Grant for Health Care (16BJZ23), CERNET Innovation Project (NGII20160701), Big Data Research & Development Project of Chinese PLA General Hospital (2018MBD-09), and National Key Research and Development Project (2017YFC0114001).

Authors' Contributions

HX and WY contributed equally to this work. HX designed the algorithm, analyzed the results, and wrote the manuscript; WY guided the data collection and data labeling; ZZ and MY provided the general guidance; KL designed and validated the algorithm; DW designed the ECG SQA algorithm; AW and ZY revised the manuscript; CM, JW, and YZ cleaned and labeled the data.

Conflicts of Interest

This work was done during ZY's internship at Beijing SensEcho Science & Technology Co, Ltd, Beijing, China, when he was a PhD candidate at University of California, Davis, CA, USA. The other authors have no conflicts to declare.

References

1. Ernsting C, Stühmann LM, Dombrowski SU, Voigt-Antons J, Kuhlmey A, Gellert P. Associations of Health App Use and Perceived Effectiveness in People With Cardiovascular Diseases and Diabetes: Population-Based Survey. JMIR Mhealth Uhealth 2019 Mar 28;7(3):e12179 [FREE Full text] [doi: 10.2196/12179] [Medline: 30920383]
2. Yang Q, Van Stee SK. The Comparative Effectiveness of Mobile Phone Interventions in Improving Health Outcomes: Meta-Analytic Review. JMIR Mhealth Uhealth 2019 Apr 03;7(4):e11244 [FREE Full text] [doi: 10.2196/11244] [Medline: 30942695]
3. World Health Organization. mHealth: new horizons for health through mobile technologies. Geneva, Switzerland: World Health Organization; 2011.
4. 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 Jul;41(7):115. [doi: 10.1007/s10916-017-0760-1] [Medline: 28631139]
5. 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;9:21 [FREE Full text] [doi: 10.1186/1743-0003-9-21] [Medline: 22520559]
6. Weenk M, van Goor H, Frietman B, Engelen LJ, van Laarhoven CJ, Smit J, et al. Continuous Monitoring of Vital Signs Using Wearable Devices on the General Ward: Pilot Study. JMIR Mhealth Uhealth 2017 Jul 05;5(7):e91 [FREE Full text] [doi: 10.2196/mhealth.7208] [Medline: 28679490]
7. Kamei T, Kanamori T, Yamamoto Y, Edirippulige S. The use of wearable devices in chronic disease management to enhance adherence and improve telehealth outcomes: A systematic review and meta-analysis. J Telemed Telecare 2020 Aug 20:1357633X20937573. [doi: 10.1177/1357633X20937573] [Medline: 32819184]
8. Ludikhuize J, Smorenburg SM, de RSE, de JE. Identification of deteriorating patients on general wards; measurement of vital parameters and potential effectiveness of the Modified Early Warning Score. J Crit Care 2012 Aug;27(4):424.e7-424.13. [doi: 10.1016/j.jcrc.2012.01.003] [Medline: 22341727]
9. Churpek MM, Yuen TC, Edelson DP. Predicting clinical deterioration in the hospital: the impact of outcome selection. Resuscitation 2013 May;84(5):564-568 [FREE Full text] [doi: 10.1016/j.resuscitation.2012.09.024] [Medline: 23022075]
10. Gambarotta N, Aletti F, Baselli G, Ferrario M. A review of methods for the signal quality assessment to improve reliability of heart rate and blood pressures derived parameters. Med Biol Eng Comput 2016 Jul;54(7):1025-1035. [doi: 10.1007/s11517-016-1453-5] [Medline: 26906277]
11. Bizzego A, Gabrieli G, Furlanello C, Esposito G. Comparison of Wearable and Clinical Devices for Acquisition of Peripheral Nervous System Signals. Sensors (Basel) 2020 Nov 27;20(23):6778 [FREE Full text] [doi: 10.3390/s20236778] [Medline: 33260880]
12. Satija U, Ramkumar B, Manikandan MS. A Review of Signal Processing Techniques for Electrocardiogram Signal Quality Assessment. IEEE Rev Biomed Eng 2018;11:36-52. [doi: 10.1109/RBME.2018.2810957] [Medline: 29994590]

13. Liu F, Wei S, Lin F, Jiang X, Liu C. An Overview of Signal Quality Indices on Dynamic ECG Signal Quality Assessment. In: Liu C, Li J, editors. Feature Engineering and Computational Intelligence in ECG Monitoring. Singapore: Springer Singapore; 2020:33-54.
14. Langley P, Marco L, King S, Duncan D, Murray A. An algorithm for assessment of quality of ECGs acquired via mobile telephones. Computing in Cardiology. 2012. URL: http://cinc.org/archives/2011/pdf/0281.pdf [accessed 2021-07-12]
15. Johannesen L. Assessment of ECG quality on an Android platform Computing in Cardiology. 2011. URL: http://cinc.org/archives/2011/pdf/0433.pdf [accessed 2021-07-12]
16. Zhang Y, Wei S, Zhang L, Liu C. Comparing the Performance of Random Forest, SVM and Their Variants for ECG Quality Assessment Combined with Nonlinear Features. J. Med. Biol. Eng 2018 Apr 27;39(3):381-392. [doi: 10.1007/s40846-018-0411-0]
17. Li Q, Mark RG, Clifford GD. Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter. Physiol Meas 2008 Jan 10;29(1):15-32 [FREE Full text] [doi: 10.1088/0967-3334/29/1/002] [Medline: 18175857]
18. Clifford G, Lopez D, Li Q, Rezek I. Signal quality indices and data fusion for determining acceptability of electrocardiograms collected in noisy ambulatory environments. Computing in Cardiology 2011;38:285-288 [FREE Full text]
19. Behar J, Oster J, Li Q, Clifford G. A single channel ECG quality metric. Computing in Cardiology 2012;39:381-384 [FREE Full text]
20. Ansari S, Gryak J, Najarian K. oise Detection in Electrocardiography Signal for Robust Heart Rate Variability Analysis: A Deep Learning Approach. New York, NY: IEEE; 2018 Presented at: 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2018; Honolulu, HI. [doi: 10.1109/embc.2018.8513537]
21. Shi Y, Han N, Li P, Yang Z, Yuan Q, Du Y. Robust Assessment of ECG Signal Quality for Wearable Devices. New York, NY: IEEE; 2019 Presented at: IEEE International Conference on Healthcare Informatics (ICHI); 2019; Xi'an, China. [doi: 10.1109/ichi.2019.8904810]
22. Satija U, Ramkumar B, Manikandan MS. An automated ECG signal quality assessment method for unsupervised diagnostic systems. Biocybernetics and Biomedical Engineering 2018;38(1):54-70. [doi: 10.1016/j.bbe.2017.10.002]
23. Zhao Z, Liu C, Li Y, Li Y, Wang J, Lin B, et al. Noise Rejection for Wearable ECGs Using Modified Frequency Slice Wavelet Transform and Convolutional Neural Networks. IEEE Access 2019;7:34060-34067. [doi: 10.1109/access.2019.2900719]
24. Charlton PH, Bonnici T, Tarassenko L, Clifton DA, Beale R, Watkinson PJ, et al. An impedance pneumography signal quality index: Design, assessment and application to respiratory rate monitoring. Biomed Signal Process Control 2021 Mar 01;65:102339. [doi: 10.1016/j.bspc.2020.102339] [Medline: 34168684]
25. Steinberg C, Philippon F, Sanchez M, Fortier-Poisson P, O'Hara G, Molin F, et al. A Novel Wearable Device for Continuous Ambulatory ECG Recording: Proof of Concept and Assessment of Signal Quality. Biosensors (Basel) 2019 Jan 21;9(1):17 [FREE Full text] [doi: 10.3390/bios9010017] [Medline: 30669678]
26. Liu C, Zhang X, Zhao L, Liu F, Chen X, Yao Y, et al. Signal Quality Assessment and Lightweight QRS Detection for Wearable ECG SmartVest System. IEEE Internet Things J 2019 Apr;6(2):1363-1374. [doi: 10.1109/jiot.2018.2844090]
27. Li L. A Quality Assessment Method of Single-Lead ECG Signal Based on Spectral Analysis. 2016 Dec 23 Presented at: 016 8th International Conference on Information Technology in Medicine and Education (ITME); 2016.12.23-2016.12.25; Fuzhou, China p. 35-38. [doi: 10.1109/ITME.2016.0018]
28. Cao D, Li D, Zhang Z, Liu X, Liang H, He M, et al. [Design and preliminary validation of a ubiquitous and wearable physiological monitoring system]. Sheng Wu Yi Xue Gong Cheng Xue Za Zhi 2019 Feb 25;36(1):121-130. [doi: 10.7507/1001-5515.201709029] [Medline: 30887786]
29. Xu H, Li P, Yang Z, Liu X, Wang Z, Yan W, et al. Construction and Application of a Medical-Grade Wireless Monitoring System for Physiological Signals at General Wards. J Med Syst 2020 Sep 04;44(10):182 [FREE Full text] [doi: 10.1007/s10916-020-01653-z] [Medline: 32885290]
30. Li Q, Clifford GD. Signal quality and data fusion for false alarm reduction in the intensive care unit. J Electrocardiol 2012 Nov;45(6):596-603. [doi: 10.1016/j.jelectrocard.2012.07.015] [Medline: 22960167]
31. Xia H, Garcia G, Mcbride J, Sullivan A, Zhao X. Computer algorithms for evaluating the quality of ECGs in real time. New York, NY: IEEE; 2011 Presented at: 2011 Computing in Cardiology; Hangzhou, China; September 18-21, 2011 p. 369-372.
32. Naseri H, Homaeinezhad M. Electrocardiogram signal quality assessment using an artificially reconstructed target lead. Comput Methods Biomech Biomed Engin 2015 Aug 27;18(10):1126-1141. [doi: 10.1080/10255842.2013.875163] [Medline: 24460414]
33. Behar J, Oster J, Qiao Li, Clifford GD. ECG Signal Quality During Arrhythmia and Its Application to False Alarm Reduction. IEEE Trans. Biomed. Eng 2013 Jun;60(6):1660-1666. [doi: 10.1109/tbme.2013.2240452]
34. Kuzilek J, Huptych M, Chudacek V, Spilka J, Lhotska L. Data driven approach to ECG signal quality assessment using multistep SVM classification. Computing in Cardiology 2011;38:453-445 [FREE Full text]
35. Zhang Y, Wei S, Long Y, Liu C. Performance Analysis of Multiscale Entropy for the Assessment of ECG Signal Quality. Journal of Electrical and Computer Engineering 2015;2015:1-9. [doi: 10.1155/2015/563915]

36. Elghazzawi Z. Method and apparatus for removing artifact from physiological signals. European Patent EP19980308473. 1999. URL: https://www.freepatentsonline.com/EP0913121.html [accessed 2021-07-12]
37. Friesen G, Jannett T, Jadallah M, Yates S, Quint S, Nagle H. A comparison of the noise sensitivity of nine QRS detection algorithms. IEEE Trans Biomed Eng 1990 Jan;37(1):85-98. [doi: 10.1109/10.43620] [Medline: 2303275]
38. Chen Y, Wu W. Isolation Forest as an Alternative Data-Driven Mineral Prospectivity Mapping Method with a Higher Data-Processing Efficiency. Nat Resour Res 2018 Feb 17;28(1):31-46. [doi: 10.1007/s11053-018-9375-6]
39. Ding Z, Fei M. An Anomaly Detection Approach Based on Isolation Forest Algorithm for Streaming Data using Sliding Window. IFAC Proceedings Volumes 2013;46(20):12-17. [doi: 10.3182/20130902-3-cn-3020.00044]
40. Liu F, Ting K, Zhou Z. Isolation Forest. 2008 Dec 15 Presented at: 2008 Eighth IEEE International Conference on Data Mining; 2008.12.15-2008.12.19; Pisa, Italy. [doi: 10.1109/ICDM.2008.17]
41. Meinard M. Dynamic time warping. In: Information retrieval for music and motion. Berlin, Heidelberg: Springer; 2007:69-84.
42. Zhang G, Kinsner W, Huang B. Electrocardiogram data mining based on frame classification by dynamic time warping matching. Comput Methods Biomech Biomed Engin 2009 Dec;12(6):701-707. [doi: 10.1080/10255840902882158] [Medline: 19360509]
43. Liu C, Zhang X, Zhao L, Liu F, Chen X, Yao Y, et al. Signal Quality Assessment and Lightweight QRS Detection for Wearable ECG SmartVest System. IEEE Internet Things J 2019 Apr;6(2):1363-1374. [doi: 10.1109/jiot.2018.2844090]
44. Kohonen T. The Basic SOM. In: Self-Organizing Maps. Berlin, Heidelberg: Springer; 1997.
45. Roy M, Gupta R, Sharma K. Photoplethysmogram Signal Quality Evaluation by Unsupervised Learning Approach. In: Photoplethysmogram Signal Quality Evaluation by Unsupervised Learning Approach. 2020 Presented at: 2020 IEEE Applied Signal Processing Conference (ASPCON); 2020.10.07-2020.10.09; Kolkata, India p. 6-10. [doi: 10.1109/ASPCON49795.2020.9276733]
46. Ghosal P, Gupta R. Classification of photoplethysmogram signal using self organizing map. In: Classification of photoplethysmogram signal using self organizing map. 2015 Presented at: 2015 IEEE International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN); 2015.11.20-2015.11.22; Kolkata, India p. 114-118. [doi: 10.1109/ICRCICN.2015.7434220]
47. Mehmood Y, Abbas M, Chen X, Honkela T. Self-Organizing Maps of Nutrition, Lifestyle and Health Situation in the World. 2011 Presented at: Advances in Self-Organizing Maps; 2011; Berlin, Heidelberg p. 160-167. [doi: 10.1007/978-3-642-21566-7_16]
48. Tibaduiza DA, Mujica LE, Rodellar J. Damage classification in structural health monitoring using principal component analysis and self-organizing maps. Struct. Control Health Monit 2012 Dec 13;20(10):1303-1316. [doi: 10.1002/stc.1540]

Abbreviations

APB: atrial premature beat
CFDA: China Food and Drug Administration
ECG: electrocardiogram
HBO: hyperbaric oxygen
IF: isolation forest
LR: logistic regression
mHealth: mobile health
MITDB: MIT-BIH Arrhythmia Database
PLAGH: Chinese PLA General Hospital
RF: random forest
RIP: respiratory inductive plethysmography
SOM: self-organizing maps
SQA: signal quality assessment
SQIs: signal quality indices
SVM: support vector machine
uHealth: ubiquitous health
VPB: ventricular premature beat
XGB: extreme gradient boosting

Edited by L Buis; submitted 31.10.20; peer-reviewed by D Li, A Tarekegn, E Park; comments to author 29.12.20; revised version received 24.01.21; accepted 25.06.21; published 12.08.21.

Please cite as:
Xu H, Yan W, Lan K, Ma C, Wu D, Wu A, Yang Z, Wang J, Zang Y, Yan M, Zhang Z
Assessing Electrocardiogram and Respiratory Signal Quality of a Wearable Device (SensEcho): Semisupervised Machine Learning-Based Validation Study
JMIR Mhealth Uhealth 2021;9(8):e25415
URL: https://mhealth.jmir.org/2021/8/e25415
doi:10.2196/25415
PMID:34387554

Review

Evaluating the Validity and Utility of Wearable Technology for Continuously Monitoring Patients in a Hospital Setting: Systematic Review

Vikas Patel[1], BMSc, MD; Ani Orchanian-Cheff[2], BA, MISt; Robert Wu[1,3], MSc, MD, FRCPC

[1]Faculty of Medicine, University of Toronto, Toronto, ON, Canada

[2]Library and Information Services, University Health Network, Toronto, ON, Canada

[3]Division of General Internal Medicine, University Health Network, Toronto, ON, Canada

Corresponding Author:
Vikas Patel, BMSc, MD
Faculty of Medicine
University of Toronto
1 King's College Cir
Toronto, ON, M5S 1A8
Canada
Phone: 1 4169756585
Email: vik.patel@mail.utoronto.ca

Abstract

Background: The term *posthospital syndrome* has been used to describe the condition in which older patients are transiently frail after hospitalization and have a high chance of readmission. Since low activity and poor sleep during hospital stay may contribute to posthospital syndrome, the continuous monitoring of such parameters by using affordable wearables may help to reduce the prevalence of this syndrome. Although there have been systematic reviews of wearables for physical activity monitoring in hospital settings, there are limited data on the use of wearables for measuring other health variables in hospitalized patients.

Objective: This systematic review aimed to evaluate the validity and utility of wearable devices for monitoring hospitalized patients.

Methods: This review involved a comprehensive search of 7 databases and included articles that met the following criteria: inpatients must be aged >18 years, the wearable devices studied in the articles must be used to continuously monitor patients, and wearables should monitor biomarkers other than solely physical activity (ie, heart rate, respiratory rate, blood pressure, etc). Only English-language studies were included. From each study, we extracted basic demographic information along with the characteristics of the intervention. We assessed the risk of bias for studies that validated their wearable readings by using a modification of the Consensus-Based Standards for the Selection of Health Status Measurement Instruments.

Results: Of the 2012 articles that were screened, 14 studies met the selection criteria. All included articles were observational in design. In total, 9 different commercial wearables for various body locations were examined in this review. The devices collectively measured 7 different health parameters across all studies (heart rate, sleep duration, respiratory rate, oxygen saturation, skin temperature, blood pressure, and fall risk). Only 6 studies validated their results against a reference device or standard. There was a considerable risk of bias in these studies due to the low number of patients in most of the studies (4/6, 67%). Many studies that validated their results found that certain variables were inaccurate and had wide limits of agreement. Heart rate and sleep were the parameters with the most evidence for being valid for in-hospital monitoring. Overall, the mean patient completion rate across all 14 studies was >90%.

Conclusions: The included studies suggested that wearable devices show promise for monitoring the heart rate and sleep of patients in hospitals. Many devices were not validated in inpatient settings, and the readings from most of the devices that were validated in such settings had wide limits of agreement when compared to gold standards. Even some medical-grade devices were found to perform poorly in inpatient settings. Further research is needed to determine the accuracy of hospitalized patients' digital biomarker readings and eventually determine whether these wearable devices improve health outcomes.

(JMIR Mhealth Uhealth 2021;9(8):e17411) doi:10.2196/17411

KEYWORDS

wearable; inpatient; continuous monitoring

Introduction

Background

Most physiologic parameters, such as vital signs or activity, are routinely monitored a few times each day in hospital ward settings [1]. Some parameters, such as sleep, are not routinely monitored at all [2,3]. More frequent monitoring could allow for the timely identification of the deteriorating health of patients and spur efforts for improving patients' overall health through increased sleep and activity. Since subtle changes in vital signs are often present 8 to 24 hours before a life-threatening event, such as intensive care unit admission or cardiac arrest, vital sign surveillance has the potential to detect clinical deterioration at an earlier phase, thereby permitting clinicians to make corrective interventions [4-7]. This includes identifying patients with poorly controlled pain and recognizing arrhythmias. The term *posthospital syndrome* has been used to denote the deleterious effects of acute illnesses that are compounded with poor sleep and low activity and occur during hospital stay [8]. Measuring sleep and activity could improve the recognition of such issues and encourage health providers to introduce interventions that improve patients' experiences in hospitals by encouraging mobilization and to identify targets for sleep-promoting interventions [9-11]. In addition, access to other digital biomarkers (eg, heart rate, blood pressure, oxygen saturation, etc) would allow clinicians to determine underlying etiologies and make tailored interventions.

The rapid uptake of affordable wearables, such as fitness bands, may provide a method for continuously measuring sleep; activity; and vital signs, such as heart rate [12-15]. However, existing literature that describes wearable devices is mostly limited to ambulatory settings and focuses on the management of chronic diseases [16,17]. More inpatient data are needed on both the validity of wearables and patient adherence. Although wearable testing has been conducted with healthy volunteers, it will be important to validate these signals in inpatient settings, where algorithms for processing sensor data into digital signals, such as those for sleep, heart rate, and activity, may be less accurate [18]. Despite the proposed benefit of intensive monitoring, many wearable studies have found issues with patient adherence [18-20]. Adherence is a crucial barrier to acquiring data and can be influenced by device convenience, the comfort of use, and interaction requirements [19]. Studies of wearable devices worn by hospitalized inpatients have been limited by large dropout rates [20].

Although there have been systematic reviews of the monitoring of patients' physical activity in hospitals [21-23], there are no reviews of the use of wearables that can reliably measure other health parameters. Therefore, in this review, we aimed to expand our search by including articles that used wearables to assess parameters other than physical activity and to assess the adherence of patients in inpatient settings.

Objective

For the purposes of this review, a wearable was considered to be any electronic device that has at least 1 sensor and can be worn on the body [24]. Wearables were examined for their ability to measure digital biomarkers, which are defined as digitally collected physiological and behavioral measures (eg, heart rate, average sleep duration, and daily step count) that explain, influence, or predict health-related outcomes [18]. Consistent with previous research, patient adherence was objectively assessed by reporting the mean proportion of patients who completed a given study [25]. The primary objectives of this review were to determine patients' adherence to using wearable devices in hospitals and to examine the validity of wearable-derived biomarker readings.

Methods

Identification and Selection of Studies

A comprehensive search strategy was developed to identify articles on the three main concepts of our question—wearables, monitoring, and inpatients. The initial search strategy was developed for Ovid MEDLINE by using a combination of database-specific subject headings and text words (Multimedia Appendix 1). Additional key words were generated based on input from the subject specialists on the team, and the revised search strategy was customized for each database.

Searches of the following databases were executed on August 16, 2018: Ovid MEDLINE, Ovid MEDLINE Epub Ahead of Print and In-Process & Other Non-Indexed Citations, Cochrane Database of Systematic Reviews, Cochrane Central Register of Controlled Trials, Health Technology Assessment database (Ovid), and CINAHL with Full Text. The search in Ovid Embase was not executed until September 5, 2018, due to issues with the vendor's August database reload. Additional search methods included reviewing the cited references of eligible studies via Web of Science (May 6, 2019) and the reference lists of eligible studies. There were no restrictions on publication period. Limits were imposed to ensure that only English-language studies and those with adult populations were included in this review. No other limits were applied to the literature search.

Article Selection and Exclusion Criteria

Records were screened by two reviewers (VP and RW) independently. For selected studies, full-text articles were obtained and evaluated for eligibility [26]. The eligibility criteria for inclusion in this review were as follows:

- Medical or surgical inpatients aged >18 years
- Device studied in the article must be a wearable (such as a watch, vest, pendant, jewelry, headset, and wristband)
- Articles must describe an element of continuous monitoring for at least 24 hours or greater
- Articles must describe the measurement of 1 or more digital biomarkers other than just physical activity or standard hospital telemetry for heart rate recording.

We excluded articles that were not considered original research, such as letters to the editor, comments, and reviews. We also excluded articles that monitored less than 3 patients, described the monitoring of a very specialized system in the body (eg, insole devices, ventricular assistive devices, and cochlear implants), involved the monitoring of patients in rehabilitation hospitals, or used wearables as tools for therapy (eg, insulin delivery).

Data Extraction

Two reviewers (RW and VP) independently extracted the data and resolved any disagreements by discussing the findings and making a collective decision. The data extracted for each article included the year of publication, study setting and design, number of participants, gender ratio, mean age of participants, digital biomarkers measured in the study, average and maximum duration that the wearable was worn by participants in each study, and patient completion rate (the proportion of patients that wore the wearable for the minimum monitoring duration that was set by the study authors). For studies that used a reference standard, any participants who were missing data from the wearable or the standard were determined to be incomplete measurement pairs and were omitted from the final count of patients who completed the study. Furthermore, we extracted the types of wearables that were worn by the participants in each study along with the placement sites on the body. Devices were classified as medical grade (approved or cleared by the US Food and Drug Administration), research grade (typically used in research settings only), and consumer grade (used by general consumers).

Validation data were also collected for each article by assessing whether the authors compared the accuracy of their digital readings to a reference standard. To determine the validity of measures that were compared to a reference standard, correlation coefficients, mean differences, and limits of agreement were extracted from each study.

Risk of Bias Assessment

All articles that assessed for validated readings were independently assessed for their risk of bias by two independent reviewers (VP and RW) using a modification of the validation subscale from a checklist for assessing the methodological quality of studies on the measurement properties of health status measurement instruments (Consensus-Based Standards for the Selection of Health Status Measurement Instruments [COSMIN]) [27] (Table 1). All discrepancies were resolved by discussion and consensus. The quality evaluation included 5 study design and methodology components (the percentage of missing data, missing data management, adequate sample size, acceptable criterion comparison, and design or methodological flaws) and 1 analysis component (acceptable accuracy analyses). We rated the quality of each dimension as excellent, good, fair, or poor based on a priori modifications to the COSMIN validation subscale for scoring criteria that are appropriate for accuracy studies (Multimedia Appendix 2) [28].

Table 1. Risk of bias assessment for studies that validated their wearable readings.

Study	Assessment criterion									
	Mean or % difference	Correlation	LOA[a]	Percentage of missing data	Missing data management	Adequate sample size (patients)	Adequate sample size (measurements)	Acceptable reference comparison	Other methodological flaws	Acceptable accuracy analyses
Bloch et al [29]	No	No	No	Excellent	Excellent	Poor	Poor	Excellent	No	Poor
Breteler et al [30]	Yes	No	Yes	Excellent	Excellent	Poor	Excellent	Excellent	No	Excellent
Gallo and Lee [13]	No	Yes	No	Excellent	Excellent	Fair	Fair	Fair	No	Excellent
Kroll et al [11,31]	Yes	Yes	Yes	Excellent	Excellent	Good	Excellent	Excellent	No	Excellent
Steinhubl et al [32]	No	Yes	No	Excellent	Excellent	Poor	Excellent	Excellent	No	Excellent
Weenk et al [4]	Yes	No	Yes	Excellent	Excellent	Poor	Good	Excellent	No	Excellent

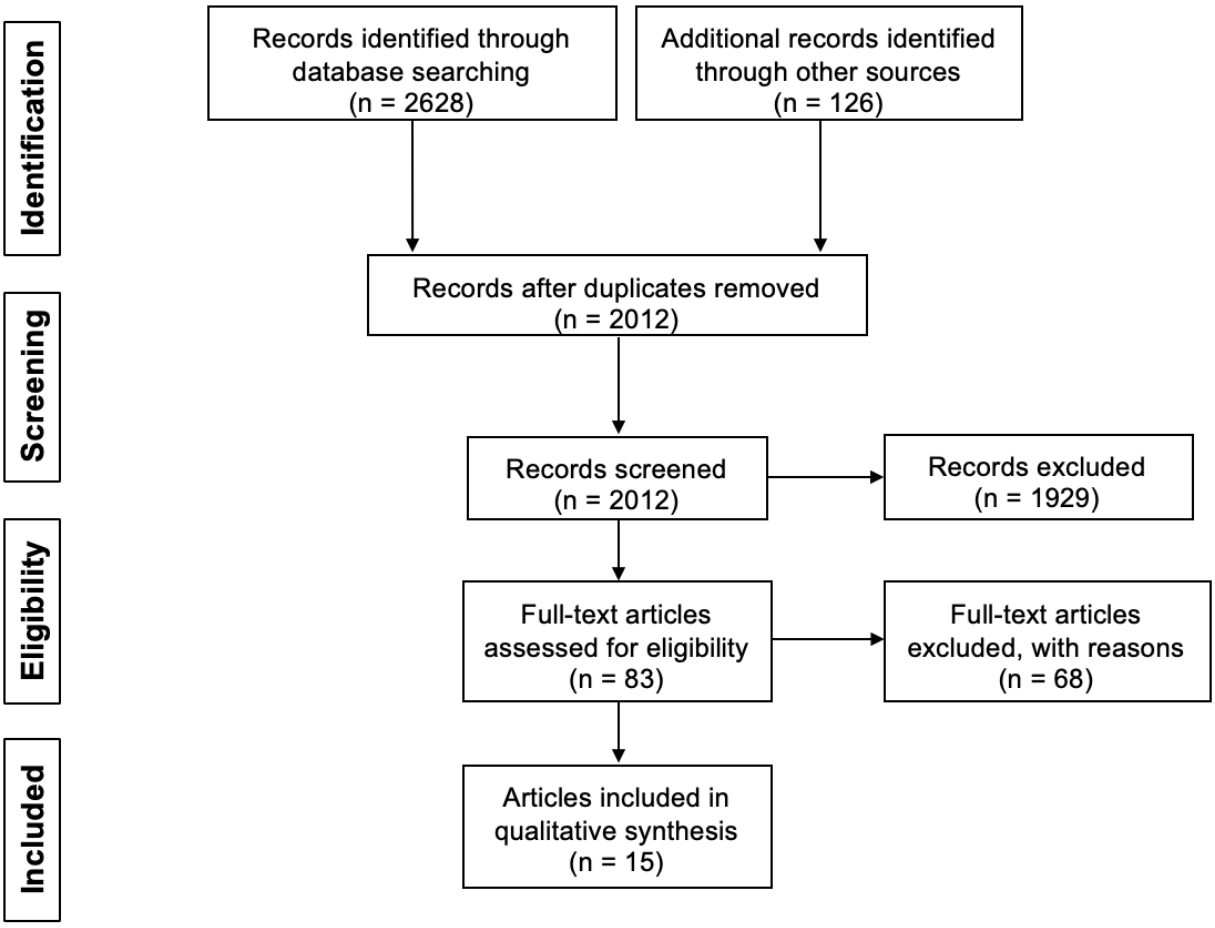
[a]LOA: limits of agreement.

Results

Characteristics of Included Studies

Our literature search identified 2754 article citations. After excluding duplicate records, 2012 records were deemed eligible for screening. A total of 83 studies were selected based on abstracts and underwent full-text review. After applying our inclusion and exclusion criteria, 15 articles that described 14 studies were selected for this review (Figure 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of included and excluded studies.



All of the articles included were prospective cohort studies (Table 2) [4,11,13,14,20,29-38]. Overall, 9 different types of commercial wearables were described across the 14 studies, and 7 different health variables were assessed collectively by the 14 studies (Table 3). The wearable devices that were described by the studies came in various different forms and were attached to a range of sites on the body (Figure 2). A total of 13 articles included both men and women as the study participants; the other two papers assessed sleep changes in postpartum women [13,34]. Kroll et al [11,31] published two articles from the same study. Both articles analyzed different aspects of the continuous monitoring of inpatients (ie, they used the same cohort of patients) but were included as the same study entry in this review (Table 2).

Collectively, the mean patient completion rate across all 14 studies was over 90%. Of the 8 articles that included a qualitative analysis as a part of their methodology, 7 reported that wearables were well received by either or both patients and clinicians.

Of the 14 studies, 6 validated wearable measurements against another standard device or measure (Table 3). The studies conducted by Bloch et al [29], Gallo and Lee [13], and Steinhubl et al [32] used intermittent measurements (nurse or questionnaires) for their reference standard. Further, Breteler et al [30] used a continuous reference (continuous electrocardiography and impedance pneumography) to compare the wearable readings for heart rates and respiratory rates [30]. Weenk et al [4] and Kroll et al [11,31] validated their wearable readings against both intermittent and continuous reference measurements. Of the 9 wearables included in the studies, 6 were cleared or approved by the US Food and Drug Administration as medical devices (ViSi Mobile [Sotera Wireless], Hidalgo EQ02 [Equivital], wrist actigraphy [Ambulatory Monitoring Inc; Actigraph LLC], LifeTouch [Isansys Lifecare], Zephyr Biopatch [Medtronic], and HealthPatch [VitalConnect]).

Table 2. Summary of included studies.

Study	Year published	Setting (ward)	Methodology	Patients, N	Male %: Female % ratio, mean age (years)	Variables measured	Number of days device was worn, average (maximum)	Patient completion rate, %
Lee and Lee [34]	2007	Obstetric	Prospective cohort	21	Females only, 32	Sleep	2[a]	100
Gallo and Lee [13]	2008	Obstetric	Prospective cohort	39	Females only, 29	Sleep	2 (2)	100
Bloch et al [29]	2011	Geriatric	Prospective cohort	10	Males and females[b], 83	Falls	21[a]	90
Chiu et al [33]	2013	Neurosurgery	Prospective cohort	60	65:35, 35	Sleep	7[a]	87
Watkins et al [37]	2015	Medicine and surgical	Prospective cohort	236	Males and females[b,c]	HR[d], RR[e], SpO_2[f], and BP[g]	3 (3)	100
Jeffs et al [20]	2016	Medicine	Prospective cohort	208	72:28[c]	HR. RR, SpO_2, temperature, and accelerometry	(14)[h]	32
Steinhubl et al [32]	2016	Medicine	Prospective cohort	26	65:35, 33	HR, RR, and temperature	3 (3)	100
Razjouyan et al [35]	2017	Hematology and oncology	Prospective cohort	35	45:55, 55	HR and fall risk	1[a]	94
Weenk et al [4]	2017	General internal medicine and surgical	Prospective cohort	20	65:35, 50	HR, RR, BP, SpO_2, and temperature	2.5 (3)	100
Kroll et al [11,31]	2017	Intensive care unit	Prospective cohort	50	52:48, 64	HR, sleep	1[a]	96
Weller et al [36]	2017	Neurology and neurosurgery	Prospective cohort	736	54:46[c]	HR, RR, SpO_2, and BP	1.7 (9)	100
Breteler et al [30]	2018	Surgical	Prospective cohort	33	72:28, 63	HR and RR	2.6 (3)	76
Yang et al [14]	2018	Oncology	Prospective cohort	11	64:36[c]	Sleep	16[a]	91
Duus et al [38]	2018	General surgery	Prospective cohort	50	58:42, 71	HR, RR, and SpO_2	3.1 (4)	100

[a]The maximum number of days was not reported in the study.

[b]The study included both male and female participants but did not report a ratio.

[c]Mean age was not reported in the study.

[d]HR: heart rate.

[e]RR: respiratory rate.

[f]SpO_2: oxygen saturation

[g]BP: blood pressure.

[h]The average number of days was not reported in the study.

Table 3. Distribution of the health variables that were assessed for accuracy in each study.

Study	Device characteristics		Digital biomarkers						
	Device, manufacturer	FDA[a] clearance or approval	Heart rate	Sleep	Respiratory rate	SpO_2[b]	Skin temperature	Blood pressure	Fall risk
Gallo and Lee [13]	Wrist Actigraph, Ambulatory Monitoring Inc	Yes	—[c]	*R*=0.53	—	—	—	—	—
Lee and Lee [34]	Mini-Motionlogger-Actigraphy, Ambulatory Monitoring Inc	Yes	—	Not validated	—	—	—	—	—
Chiu et al [33]	ActiGraph GT1M, Actigraph LLC	Yes	—	Not validated	—	—	—	—	—
Yang et al [14]	Actigraph GT3X+ watch, Actigraph LLC	Yes	—	Not validated	—	—	—	—	—
Kroll et al [11,31]	Fitbit Charge HR, Fitbit Inc	—	LoA[d] (sinus): 23.9 to 21.9 beats per minute	*R*=0.33	—	—	—	—	—
Breteler et al [30]	HealthPatch, VitalConnect	Yes	LoA: −8.8 to 6.5 beats per minute	—	LoA: −15.8 to 11.2 breaths per minute	—	Not validated	—	Not validated
Jeffs et al [20]	Hidalgo EQ02, Equivital	Yes	Not validated	—	Not validated	Not validated	Not validated	—	—
Duus et al [38]	LifeTouch, Isansys Lifecare	Yes	Not validated	—	Not validated	Not validated	—	—	—
Steinhubl et al [32][e]	MultiSense patch, Rhythm Diagnostic Systems	—	*R*=0.75	—	*R*=0.83	—	*R*=0.99	—	—
Bloch et al [29]	Vigi'Fall, Vigilio Telemedical	—	—	—	—	—	—	—	Sensitivity: 37.5%
Weenk et al [4]	ViSi Mobile, Sotera Wireless	Yes	LoA: −11.1 to 10.7 beats per minute	—	−5.5 to 7.9 breaths per minute	−3.1% to 3.3%	Not validated	SBP[f]: −23 to 24 mm Hg; DBP[g]: 27.5 to 11.5 mm Hg	—
Weenk et al [4]	HealthPatch, VitalConnect	Yes	−12.6 to 9.5 beats per minute	—	−10.3 to 9.0 breaths per minute	Not validated	Not validated	Not validated	—
Weller et al [36]	ViSi Mobile, Sotera Wireless	Yes	Not validated	—	Not validated	Not validated	Not validated	Not validated	—
Watkins et al [37]	ViSi Mobile, Sotera Wireless	Yes	Not validated	—	Not validated	Not validated	—	Not validated	—
Razjouyan et al [35]	Zephyr BioPatch, Medtronic	Yes	Not validated	—	—	—	—	—	Not validated

[a]FDA: US Food and Drug Administration.

[b]SpO_2: oxygen saturation.

[c]Not available.

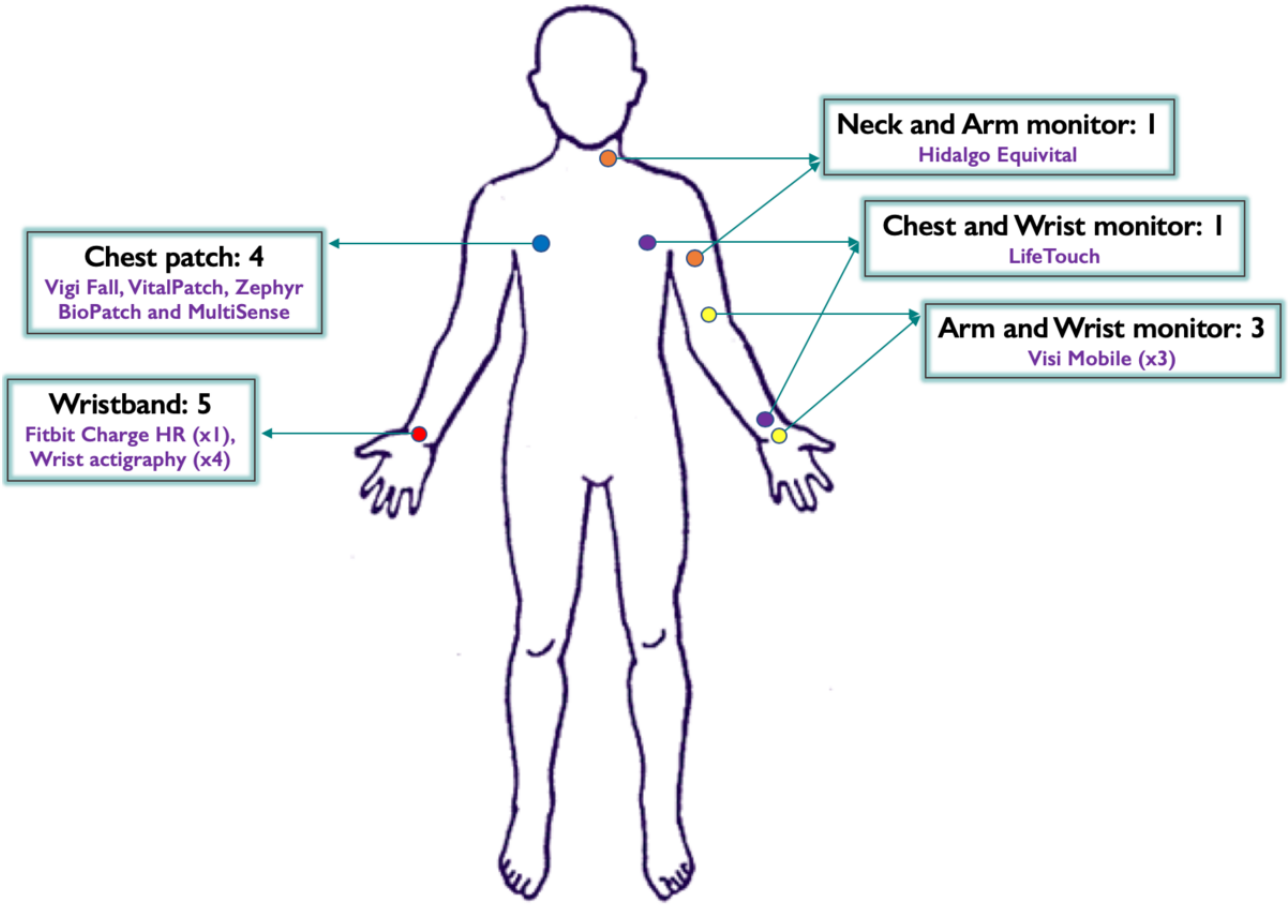
[d]LOA: limits of agreement.

[e]Steinhubl et al [32] did not report limits of agreement.

[f]SBP: systolic blood pressure.

[g]DBP: diastolic blood pressure.

Figure 2. Illustration of the types of and body locations for used wearable devices.



Risk of Bias

Of the 6 studies in the risk of bias assessment, 4 were ranked as poor due to a small sample size (participants: N<30). The study conducted by Gallo and Lee [13] used sleep questionnaires as a reference measure and therefore received a fair rating for the "acceptability of reference" criterion, whereas the other five studies were ranked as excellent (ie, they used intermittent nurse readings or other validated methodologies). Further, in terms of assessing the accuracy analyses, only the study conducted by Bloch et al [29] did not report mean differences, correlations, and limits of agreement.

Validation by Digital Biomarker

Heart Rate

A total of 5 studies assessed heart rate accuracy. Breteler et al [30] found that the bias and 95% limits of agreement for heart rate were −1.1 beats per minute (BPM) and −8.8 to 6.5 BPM, respectively, for 55,565 heart rate pairs [30]. Specifically, the wearable sensor accurately detected tachycardia with a sensitivity of 90% and a specificity of 97% [30]. In a cohort of intensive care unit patients, Kroll et al [11,31] found that the Fitbit (Fitbit Inc)-derived heart rate values were slightly lower than those derived from continuous electrocardiography monitoring but that 73% of the readings were within 5 BPM of the electrocardiogram value (average bias: −1.14 BPM; R=0.74; P<.001; heart rate pairs: n=12,358) [11]. Overall, the limit of agreement for the Fitbit device was 24 BPM, but its performance was significantly better in patients in sinus rhythm than in those who were not in sinus rhythm (average bias: −0.99 BPM vs −5.02 BPM, respectively; P=.02; limits of agreement: 22.9 BPM vs 46.4 BMP, respectively; P=.049) [11]. Kroll et al [11,31] also found that the Fitbit was very specific when it detected tachycardia (sensitivity=70%; specificity=99%) [31]. Steinhubl et al [32] demonstrated that manual and automated heart rate readings correlated well (R=0.75; measurements: n=111), but limits of agreement were not reported [32]. Weenk et al [4] reported that heart rate readings were generally consistent when compared to the nurse recordings; the limits of agreement for the ViSi Mobile and the HealthPatch were −11.1 to 10.7 BPM and −12.6 to 9.5 BPM, respectively (86 measurements).

Sleep

A total of 6 studies used wearables to assess sleep, of which 2 assessed whether wearable readings were reliable. Gallo and Lee [13] found that self-reported sleep correlated with the actigraphy-recorded number of awakenings (R=0.53; P=.01) [13]. Kroll et al [11,31] found that there was a moderate correlation between wearable-derived sleep duration and questionnaire-derived sleep quality (R=0.33; P=.03) [31].

Respiratory Rate

Of the 8 articles that used different wearables to measure the respiratory rate of patients, 3 assessed the wearables' accuracy. Breteler et al [30] found that for respiratory rate, the bias was −2.3 breaths per minute, and wide limits of agreement were reported (−15.8 to 11.2 breaths per minute; measurement pairs: n=56,674) [30]. Steinhubl et al [32] reported that there was a strong correlation between wearable and manual respiratory rate readings (R=0.83; P<.001; measurements: n=111), but limits of agreement were not reported [32]. Weenk et al [4] described wide limits of agreement for respiratory rate based on 86 measurements (ViSi Mobile limits of agreement: −5.5 to 7.9 breaths per minute; HealthPatch limits of agreement: −10.3 to 9.0 breaths per minute) [4].

Other Measures

Only 1 study, which was conducted by Weenk et al [4], assessed the accuracy of oxygen saturation and blood pressure readings from ViSi Mobile by comparing them to HealthPatch readings as well as intermittent nurse measurements. From 86 measurements, they found that the automated readings for the systolic blood pressure, diastolic blood pressure, and oxygen saturation had wide limits of agreement (systolic blood pressure: −23.1 to 24.0 mm Hg; diastolic blood pressure: −27.5 to 11.5 mm Hg; oxygen saturation: −3.1% to 3.3%) [4]. Of the 6 articles that used wearables that measured skin temperature, only Steinhubl et al [32] validated the results against a reference standard to conclude that the automated readings were reliable (R=0.99; n=112), but bias and limits of agreement were not reported [32]. Of the 3 articles in this review that detected falls by using wearables, only Bloch and colleagues [29] assessed accuracy and found that the Vigi'Fall system had a low sensitivity (37.5%) to fall risk [29].

Discussion

Principal Findings

We conducted a systematic review that evaluated the utility of wearable technology in continuously monitoring hospitalized patients for a wide variety of health parameters. Our review focused on the breadth of devices used and the signals measured in hospitalized patients and included consumer, research, and medical-grade devices. There was evidence to support the use of Fitbit, ViSi Mobile, and the HealthPatch to measure heart rate [4,11,31], since the readings were validated against both intermittent and continuous reference standards. This review demonstrated that the validity of the data did not necessarily correlate with the classification of the device because even some medical-grade devices did not perform well and yielded data with wide limits of agreement. We found that only 6 studies validated the accuracy of wearable-derived health data from hospitalized patients by comparing the readings against a reference standard. Overall, the quality of most of these studies was excellent in terms of the reporting of missing data (6/6, 100%) and the use of acceptable accuracy evaluations (5/6, 83%). However, there was a considerable risk of bias in these studies due to the low number of participants in most of the studies (4/6, 67%). Many studies reported wide limits of agreement for other digital biomarkers, such as respiratory rate and blood pressure. Of note, we also found that the majority of studies (8/14, 57%) did not validate the studied device or parameter measured.

Of the various health parameters, the best evidence of validity was in the monitoring of heart rate in hospitalized patients. We also found that, in hospital settings, limits of agreement for medical-grade devices ranged from 16.4 to 21.8 BPM, whereas the limit for a Fitbit consumer device that uses photoplethysmography signals was 24 BPM. Further, during Fitbit-based continuous electrocardiogram monitoring, 73% of the readings were within 5 BPM of electrocardiogram readings. In a systematic review of 158 studies that measured heart rate by using consumer wearable devices, 71% and 51% of Apple Watch (Apple Inc) readings (used in 49 studies) and Fitbit readings (used in 71 studies), respectively, were within 3% of electrocardiogram readings in controlled settings [39]. Moreover, in 3 free-living studies, the wrist-worn Fitbit Charge had a mean absolute error percentage of 10% [39]. A systematic review of wrist-worn devices that measure heart rate via plethysmography found limits of agreement of 8.4 BPM at rest, 30.1 BPM while on a treadmill, and 41.5 BPM while cycling [40]. Overall, our findings found large limits of agreement for all devices, and inpatient results were consistent with the wide limits of agreement found in free-living environments or with activity.

We found that sleep only had a moderate correlation with sleep survey results from inpatient settings the use research and consumer devices. A recent systematic review of Fitbit-based sleep assessments found that readings from more recently developed devices correlated well with polysomnography readings for assessing sleep episodes [41]. It is unclear whether the lower correlation that we found was due to inpatient settings with high nighttime interruptions, patient factors that were perhaps associated with acute illness, or issues with sleep surveys (or a combination of these three factors) [3]. With respect to respiratory rate, 2 studies of 2 medical-grade devices provided limits of agreement. Wider limits of agreement were found in the study that had over 50,000 measurement pairs and used a gold standard (27 breaths per minute) compared to those in the study that had less than 100 measurement pairs and used clinician-reported vitals (13.4-19.3 breaths per minute) [30,32]. Additionally, previous studies found that medical-grade devices were only accurate under laboratory conditions or at-home conditions [42,43]. There was a limited number of studies on oxygen saturation, temperature, blood pressure, and fall risk.

Limitations and Future Research

There are a few limitations that should be noted for our systematic review. There is a considerable risk of bias, as the number of participants in the studies was low. Further, the studies included were observational in design and had a high degree of heterogeneity in terms of the objectives, populations, and outcomes reported. Thus, the data analysis methods were limited to broad categorization and the extraction of the common themes and trends that emerged from the results. Reports of wearable monitoring from individual studies should be viewed based on their methodological limitations. Although patient adherence has been found to correlate well with patients'

acceptability of wearables devices in inpatient settings, we realize that studying factors such as data loss, the duration of data gaps, and qualitative feedback from nurses and patients would further strengthen the generalizability of the results. Finally, it is important to note that wearable studies are being increasingly performed, and more relevant articles will become increasingly available.

This review also identifies gaps in knowledge that still exist within literature and provides information about what is required for further research. Specifically, the further validation of digital biomarkers by using gold standard comparators, such as polysomnography for assessing sleep and continuous electrocardiogram monitoring for assessing heart rate, is required. Ideally, large participant sample sizes and large numbers of measurement pairs within a population of interest should be used to assess parameters such as vital signs. The use of 2 reference standards to validate each health parameter, such as a heart rate, has also been recommended [44]. Moreover, data that are derived under real life conditions are still needed to better understand the factors that may contribute to between-patient heterogeneity when comparing the accuracy of wearable readings, such as those for patient activity, posture, gait type and velocity, locations of wearables, and patients' diagnoses (eg, seizures). Future studies can aim to further qualify the process of retrieving data by using wearables to explore other barriers and avenues that might hinder the collection of reliable health information (ie, a weak Bluetooth connectivity, a lack of patient digital health literacy, the added burden that the process of taking wearable readings has on clinicians, the learning curve required to operate a wearable, etc) Finally, while we found that some digital biomarkers appeared to be valid for the monitoring of inpatients via wearables, we were unable to find any studies that supported the use of wearables in inpatient settings to improve clinical outcomes.

Conclusions

Overall, the assessment of studies in this review suggested that wearable devices show promise for monitoring the heart rate and sleep of patients in hospitals. The results show that many devices were not validated in inpatient settings, and the readings from most of the devices that were validated in such settings had wide limits of agreement. Further research is needed to determine the accuracy of the digital biomarker readings of hospitalized patients and to eventually determine whether wearable devices improve the health outcomes of hospitalized patients.

Acknowledgments

RW is supported by an award from the Mak Pak Chiu and Mak-Soo Lai Hing Chair of General Internal Medicine, University of Toronto. Funding for this study was kindly provided by the University Health Network Foundation Complex Care Fund.

Authors' Contributions

VP, AOC, and RW designed and planned the review. AOC conducted the search strategy. VP and RW screened the articles and conducted the data analysis. VP, AOC, and RW wrote and revised the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search methods.
[DOCX File , 64 KB - mhealth_v9i8e17411_app1.docx]

Multimedia Appendix 2
Modified Consensus-Based Standards for the Selection of Health Status Measurement Instruments (COSMIN) criteria used for the risk of bias assessment.
[DOCX File , 30 KB - mhealth_v9i8e17411_app2.docx]

References

1. Mariani P, Saeed MU, Potti A, Hebert B, Sholes K, Lewis MJ, et al. Ineffectiveness of the measurement of 'routine' vital signs for adult inpatients with community-acquired pneumonia. Int J Nurs Pract 2006 Apr;12(2):105-109. [doi: 10.1111/j.1440-172X.2006.00556.x] [Medline: 16529596]
2. Kessler R, Knutson KL, Mokhlesi B, Anderson SL, Shah M, Meltzer DO, et al. Sleep and activity patterns in older patients discharged from the hospital. Sleep 2019 Oct 21;42(11):zsz153 [FREE Full text] [doi: 10.1093/sleep/zsz153] [Medline: 31310317]
3. Stewart NH, Arora VM. Sleep in hospitalized older adults. Sleep Med Clin 2018 Mar;13(1):127-135 [FREE Full text] [doi: 10.1016/j.jsmc.2017.09.012] [Medline: 29412979]

4. Weenk M, van Goor H, Frietman B, Engelen LJ, van Laarhoven CJ, Smit J, et al. Continuous monitoring of vital signs using wearable devices on the general ward: Pilot study. JMIR Mhealth Uhealth 2017 Jul 05;5(7):e91 [FREE Full text] [doi: 10.2196/mhealth.7208] [Medline: 28679490]
5. Mapp ID, Davis LL, Krowchuk H. Prevention of unplanned intensive care unit admissions and hospital mortality by early warning systems. Dimens Crit Care Nurs 2013;32(6):300-309. [doi: 10.1097/DCC.0000000000000004] [Medline: 24100432]
6. Churpek MM, Yuen TC, Edelson DP. Predicting clinical deterioration in the hospital: the impact of outcome selection. Resuscitation 2013 May;84(5):564-568 [FREE Full text] [doi: 10.1016/j.resuscitation.2012.09.024] [Medline: 23022075]
7. Cretikos M, Chen J, Hillman K, Bellomo R, Finfer S, Flabouris A. The objective medical emergency team activation criteria: a case-control study. Resuscitation 2007 Apr;73(1):62-72. [doi: 10.1016/j.resuscitation.2006.08.020] [Medline: 17241732]
8. Krumholz HM. Post-hospital syndrome--an acquired, transient condition of generalized risk. N Engl J Med 2013 Jan 10;368(2):100-102 [FREE Full text] [doi: 10.1056/NEJMp1212324] [Medline: 23301730]
9. Alharbi M, Straiton N, Smith S, Neubeck L, Gallagher R. Data management and wearables in older adults: A systematic review. Maturitas 2019 Jun;124:100-110. [doi: 10.1016/j.maturitas.2019.03.012] [Medline: 30910279]
10. Whyman J. Post-hospital syndrome: Tips to keep yourself or a loved one healthy after hospitalization. Harvard Health Publishing. URL: https://www.health.harvard.edu/blog/post-hospital-syndrome-tips-to-keep-yourself-or-a-loved-one-healthy-after-hospitalization-2019012315830 [accessed 2021-07-29]
11. Kroll RR, Boyd JG, Maslove DM. Accuracy of a wrist-worn wearable device for monitoring heart rates in hospital inpatients: A prospective observational study. J Med Internet Res 2016 Sep 20;18(9):e253 [FREE Full text] [doi: 10.2196/jmir.6025] [Medline: 27651304]
12. Welch J, Kanter B, Skora B, McCombie S, Henry I, McCombie D, et al. Multi-parameter vital sign database to assist in alarm optimization for general care units. J Clin Monit Comput 2016 Dec;30(6):895-900 [FREE Full text] [doi: 10.1007/s10877-015-9790-8] [Medline: 26439830]
13. Gallo A, Lee KA. Sleep characteristics in hospitalized antepartum patients. J Obstet Gynecol Neonatal Nurs 2008;37(6):715-721 [FREE Full text] [doi: 10.1111/j.1552-6909.2008.00297.x] [Medline: 19012722]
14. Yang CJ, Aibel K, Meyerhoff R, Wang F, Harpole D, Abernethy AP, et al. Actigraphy assessment of sleep quality among patients with acute myeloid leukaemia during induction chemotherapy. BMJ Support Palliat Care 2018 Sep;8(3):274-277. [doi: 10.1136/bmjspcare-2018-001509] [Medline: 29643104]
15. Mercer K, Li M, Giangregorio L, Burns C, Grindrod K. Behavior change techniques present in wearable activity trackers: A critical analysis. JMIR Mhealth Uhealth 2016;4(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4461] [Medline: 27122452]
16. Wieringa FP, Broers NJH, Kooman JP, Van Der Sande FM, Van Hoof C. Wearable sensors: can they benefit patients with chronic kidney disease? Expert Rev Med Devices 2017 Jul;14(7):505-519. [doi: 10.1080/17434440.2017.1342533] [Medline: 28612635]
17. Kirk MA, Amiri M, Pirbaglou M, Ritvo P. Wearable technology and physical activity behavior change in adults with chronic cardiometabolic disease: A systematic review and meta-analysis. Am J Health Promot 2019 Jun;33(5):778-791. [doi: 10.1177/0890117118816278] [Medline: 30586996]
18. Sim I. Mobile devices and health. N Engl J Med 2019 Sep 05;381(10):956-968. [doi: 10.1056/NEJMra1806949] [Medline: 31483966]
19. Wright SP, Brown TSH, Collier SR, Sandberg K. How consumer physical activity monitors could transform human physiology research. Am J Physiol Regul Integr Comp Physiol 2017 Mar 01;312(3):R358-R367. [doi: 10.1152/ajpregu.00349.2016] [Medline: 28052867]
20. Jeffs E, Vollam S, Young JD, Horsington L, Lynch B, Watkinson PJ. Wearable monitors for patients following discharge from an intensive care unit: practical lessons learnt from an observational study. J Adv Nurs 2016 Aug;72(8):1851-1862. [doi: 10.1111/jan.12959] [Medline: 26990704]
21. Baldwin C, van Kessel G, Phillips A, Johnston K. Accelerometry shows inpatients with acute medical or surgical conditions spend little time upright and are highly sedentary: Systematic review. Phys Ther 2017 Nov 01;97(11):1044-1065. [doi: 10.1093/ptj/pzx076] [Medline: 29077906]
22. Lim SER, Ibrahim K, Sayer AA, Roberts HC. Assessment of physical activity of hospitalised older adults: A systematic review. J Nutr Health Aging 2018;22(3):377-386. [doi: 10.1007/s12603-017-0931-2] [Medline: 29484351]
23. Anderson JL, Green AJ, Yoward LS, Hall HK. Validity and reliability of accelerometry in identification of lying, sitting, standing or purposeful activity in adult hospital inpatients recovering from acute or critical illness: a systematic review. Clin Rehabil 2018 Feb;32(2):233-242. [doi: 10.1177/0269215517724850] [Medline: 28805075]
24. Yetisen AK, Martinez-Hurtado JL, Ünal B, Khademhosseini A, Butt H. Wearables in medicine. Adv Mater 2018 Jun 11:e1706910 [FREE Full text] [doi: 10.1002/adma.201706910] [Medline: 29893068]
25. Pavic M, Klaas V, Theile G, Kraft J, Tröster G, Guckenberger M. Feasibility and usability aspects of continuous remote monitoring of health status in palliative cancer patients using wearables. Oncology 2020;98(6):386-395 [FREE Full text] [doi: 10.1159/000501433] [Medline: 31336377]
26. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097. [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]

27. 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]
28. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012 May;21(4):651-657 [FREE Full text] [doi: 10.1007/s11136-011-9960-1] [Medline: 21732199]
29. Bloch F, Gautier V, Noury N, Lundy J, Poujaud J, Claessens Y, et al. Evaluation under real-life conditions of a stand-alone fall detector for the elderly subjects. Ann Phys Rehabil Med 2011 Sep;54(6):391-398 [FREE Full text] [doi: 10.1016/j.rehab.2011.07.962] [Medline: 21903502]
30. Breteler MJM, Huizinga E, van Loon K, Leenen LPH, Dohmen DAJ, Kalkman CJ, et al. Reliability of wireless monitoring using a wearable patch sensor in high-risk surgical patients at a step-down unit in the Netherlands: a clinical validation study. BMJ Open 2018 Dec 27;8(2):e020162 [FREE Full text] [doi: 10.1136/bmjopen-2017-020162] [Medline: 29487076]
31. Kroll RR, McKenzie ED, Boyd JG, Sheth P, Howes D, Wood M, WEARable Information Technology for hospital INpatients (WEARIT-IN) study group. Use of wearable devices for post-discharge monitoring of ICU patients: a feasibility study. J Intensive Care 2017;5:64 [FREE Full text] [doi: 10.1186/s40560-017-0261-9] [Medline: 29201377]
32. Steinhubl SR, Feye D, Levine AC, Conkright C, Wegerich SW, Conkright G. Validation of a portable, deployable system for continuous vital sign monitoring using a multiparametric wearable sensor and personalised analytics in an Ebola treatment centre. BMJ Glob Health 2016 Jul 05;1(1):e000070 [FREE Full text] [doi: 10.1136/bmjgh-2016-000070] [Medline: 28588930]
33. Chiu H, Lo W, Chiang Y, Tsai P. The effects of sleep on the relationship between brain injury severity and recovery of cognitive function: a prospective study. Int J Nurs Stud 2014 Jun;51(6):892-899. [doi: 10.1016/j.ijnurstu.2013.10.020] [Medline: 24246095]
34. Lee SY, Lee KA. Early postpartum sleep and fatigue for mothers after cesarean delivery compared with vaginal delivery: an exploratory study. J Perinat Neonatal Nurs 2007;21(2):109-113. [doi: 10.1097/01.jpn.0000270627.73993.b0]
35. Razjouyan J, Grewal GS, Rishel C, Parthasarathy S, Mohler J, Najafi B. Activity monitoring and heart rate variability as indicators of fall risk: Proof-of-concept for application of wearable sensors in the acute care setting. J Gerontol Nurs 2017 Jul 01;43(7):53-62. [doi: 10.3928/00989134-20170223-01] [Medline: 28253410]
36. Weller RS, Foard KL, Harwood TN. Evaluation of a wireless, portable, wearable multi-parameter vital signs monitor in hospitalized neurological and neurosurgical patients. J Clin Monit Comput 2018 Oct;32(5):945-951. [doi: 10.1007/s10877-017-0085-0] [Medline: 29214598]
37. Watkins T, Whisman L, Booker P. Nursing assessment of continuous vital sign surveillance to improve patient safety on the medical/surgical unit. J Clin Nurs 2016 Jan;25(1-2):278-281 [FREE Full text] [doi: 10.1111/jocn.13102] [Medline: 26531215]
38. Duus CL, Aasvang EK, Olsen RM, Sørensen HBD, Jørgensen LN, Achiam MP, et al. Continuous vital sign monitoring after major abdominal surgery-Quantification of micro events. Acta Anaesthesiol Scand 2018 Oct;62(9):1200-1208. [doi: 10.1111/aas.13173] [Medline: 29963706]
39. 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]
40. Zhang Y, Weaver RG, Armstrong B, Burkart S, Zhang S, Beets MW. Validity of wrist-worn photoplethysmography devices to measure heart rate: A systematic review and meta-analysis. J Sports Sci 2020 Sep;38(17):2021-2034. [doi: 10.1080/02640414.2020.1767348] [Medline: 32552580]
41. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Accuracy of wristband Fitbit models in assessing sleep: Systematic review and meta-analysis. J Med Internet Res 2019 Nov 28;21(11):e16273 [FREE Full text] [doi: 10.2196/16273] [Medline: 31778122]
42. Chan AM, Selvaraj N, Ferdosi N, Narasimhan R. Wireless patch sensor for remote monitoring of heart rate, respiration, activity, and falls. 2013 Presented at: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); July 3-7, 2013; Osaka, Japan. [doi: 10.1109/embc.2013.6610948]
43. Chan AM, Ferdosi N, Narasimhan R. Ambulatory respiratory rate detection using ECG and a triaxial accelerometer. 2013 Presented at: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); July 3-7, 2013; Osaka, Japan. [doi: 10.1109/embc.2013.6610436]
44. Sartor F, Papini G, Cox LGE, Cleland J. Methodological shortcomings of wrist-worn heart rate monitors validations. J Med Internet Res 2018 Dec 02;20(7):e10108 [FREE Full text] [doi: 10.2196/10108] [Medline: 29967000]

Abbreviations

BPM: beats per minute
COSMIN: Consensus-Based Standards for the Selection of Health Status Measurement Instruments

Edited by L Buis; submitted 12.12.19; peer-reviewed by M Breteler, D Atsma, L Li, K Ng, R Eckhoff; comments to author 28.08.20; revised version received 21.10.20; accepted 15.07.21; published 18.08.21.

Please cite as:
Patel V, Orchanian-Cheff A, Wu R
Evaluating the Validity and Utility of Wearable Technology for Continuously Monitoring Patients in a Hospital Setting: Systematic Review
JMIR Mhealth Uhealth 2021;9(8):e17411
URL: https://mhealth.jmir.org/2021/8/e17411
doi:10.2196/17411
PMID:34406121

Original Paper

Prosociality and the Uptake of COVID-19 Contact Tracing Apps: Survey Analysis of Intergenerational Differences in Japan

Masahiro Shoji[1*], PhD; Asei Ito[1*], PhD; Susumu Cato[1*], PhD; Takashi Iida[1*], LLM; Kenji Ishida[1*], PhD; Hiroto Katsumata[2*], PhD; Kenneth Mori McElwain[1*], PhD

[1]Institute of Social Science, University of Tokyo, Tokyo, Japan

[2]Graduate School of Arts and Sciences, University of Tokyo, Tokyo, Japan

[*]all authors contributed equally

Corresponding Author:
Masahiro Shoji, PhD
Institute of Social Science
University of Tokyo
7-3-1 Hongo, Bunkyo-ku
Tokyo, 113-0033
Japan
Phone: 81 3 5841 4927
Email: shoji@iss.u-tokyo.ac.jp

Abstract

Background: To control the COVID-19 pandemic, it is essential to trace and contain infection chains; for this reason, policymakers have endorsed the usage of contact tracing apps. To date, over 50 countries have released such apps officially or semiofficially, but those that rely on citizens' voluntary uptake suffer from low adoption rates, reducing their effectiveness. Early studies suggest that the low uptake is driven by citizens' concerns about security and privacy, as well as low perceptions of infection risk and benefits from the usage. However, these do not explore important generational differences in uptake decision or the association between individuals' prosociality and uptake.

Objective: The objective of our study was to examine the role of individuals' prosociality and other factors discussed in the literature, such as perceived risk and trust in government, in encouraging the usage of contact tracing apps in Japan. We paid particular attention to generational differences.

Methods: A web-based survey was conducted in Japan 6 months after the release of a government-sponsored contact tracing app. Participants were recruited from individuals aged between 20 and 69 years. Exploratory factor analyses were conducted to measure prosociality, risk perception, and trust in government. Logistic regression was used to examine the association between these factors and uptake.

Results: There was a total of 7084 respondents, and observations from 5402 respondents were used for analysis, of which 791 respondents (14.6%) had ever used the app. Two factors of prosociality were retained: agreeableness and attachment to the community. Full-sample analysis demonstrated app uptake was determined by agreeableness, attachment to the community, concern about health risks, concern about social risks, and trust in the national government; however, important differences existed. The uptake decision of respondents aged between 20 and 39 years was attributed to their attachment to the community (odds ratio [OR] 1.28, 95% CI 1.11-1.48). Agreeable personality (OR 1.18, 95% CI 1.02-1.35), concern about social risk (OR 1.17, 95% CI 1.02-1.35), and trust in national government (OR 1.16, 95% CI 1.05-1.28) were key determinants for those aged between 40 and 59 years. For those aged over 60 years, concerns about health risks determined the uptake decision (OR 1.49, 95% CI 1.24-1.80).

Conclusions: Policymakers should implement different interventions for each generation to increase the adoption rate of contact tracing apps. It may be effective to inform older adults about the health benefits of the apps. For middle-age adults, it is important to mitigate concerns about security and privacy issues, and for younger generations, it is necessary to boost their attachment to their community by utilizing social media and other web-based network tools.

(JMIR Mhealth Uhealth 2021;9(8):e29923) doi:10.2196/29923

KEYWORDS

COVID-19; contact tracing app; place attachment; place identity; contact tracing; pandemic; mHealth; health policy

Introduction

Background

The COVID-19 pandemic has caused immense human and socioeconomic harm worldwide [1,2]. To contain the pandemic, it is essential to track infection chains and prevent further spread of infections. Traditionally, this was performed manually through call centers, but given the progress of digital technology and mobile phones, policymakers have increasingly endorsed the usage of contact tracing apps [3]. These apps send users a warning message when other users with whom they have been in close contact are confirmed to be infected. As of January 2021, over 50 nations have released such apps officially or semiofficially [4], and policymakers expected this new technology to play a pivotal role in controlling the infection spread [5].

However, to date, the apps have not been as successful as originally expected in many countries. According to an early simulation, a 56% adoption rate was necessary to contain the virus effectively [6]. The governments of Singapore and Qatar required that citizens download contact tracing apps and achieved over 80% adoption. By contrast, many countries that have relied on citizens' voluntary uptake have failed to reach the required level of uptake. In the United Kingdom, only 28.5% of citizens installed the app, and in Germany, the download and installation rate was as low as 21.7% at the end of January 2021 [4].

Studies [7-11] have suggested that low uptake rates were driven by citizens' concerns about security and privacy, poor trust in government, low perceived infection risk, and low perceived benefit from usage; however, 2 issues remained unaddressed. First, generational differences in uptake decisions were largely unexplored. This issue is crucial given differences in the health impact of infections and in familiarity with mobile apps across age groups. Second, little attention was paid to an important characteristic of the apps—the apps prevent users from spreading the virus but do not protect users themselves from infections. Therefore, motives for using the app could vary among individuals by their prosociality or their willingness to engage in prosocial behavior. Prosocial behavior is "a broad category of actions that are defined by society as generally beneficial to other people and to the ongoing political system [12]." The prosocial feature of contact tracing apps is crucial in the context of the COVID-19 pandemic, because early studies [13] warned that economic losses and lack of social interactions due to mandatory or voluntary social distancing may aggravate individuals' antisocial behavior. It is also important to discuss whether disaster-prevention infrastructure and community play complementary or substitute roles in disaster preparedness, response, and resilience; these are major issues in disaster research [14].

Aims

The goal of this study was to examine the role of individuals' prosociality and other factors discussed in the literature, such as perceived risk and trust in government, in encouraging the usage of contact tracing apps in Japan. We paid particular attention to generational differences. Uncovering generational differences will enable policymakers to tailor interventions to each age group. This is relevant in countries where specific age ranges are exposed to higher infection risks than others. In Japan, 744,953 people were confirmed to be infected, of which 37% were between 20 and 39 years of age (as of June 2, 2021) [15]. Controlling the spread of infection among this generation is especially important.

Methods

Survey Design

We conducted an original nationwide web-based survey in Japan, which was designed to collect data from approximately 7000 people aged between 20 and 69 years. In the sampling process, 68,480 people were selected from registrants of a large (4.65 million registrants) survey company in Japan (Cross Marketing Inc). Registrants were randomly sampled with stratification with respect to gender (2 categories), age (10 categories with 5-year ranges), and location of residence (10 categories: Hokkaido, Tohoku, Minami-Kanto, Kita-Kanto and Koshin, Hokuriku, Tokai, Kinki, Chugoku, Shikoku, and Kyushu), so that the expected distribution of these characteristics was comparable to that of the Japanese population.

The Japanese government released its official mobile app—the COVID-19 Contact Confirming App (COCOA)—in June 2020. The invitation for the survey was sent to 68,480 members by email on December 18, 2020, 6 months after the release of COCOA. Participants were informed that they would receive shopping tokens as a financial incentive and that the survey would be closed once the required sample size was obtained. The survey was closed on December 21, by which time 9369 individuals read the informed consent on the survey website, 7997 agreed to participate, and 7084 completed the survey (response rate among those who visited the website: 75.6%). The questionnaire collected information about respondents, including protective measures taken against COVID-19, perceived risks from COVID-19 pandemic, the usage of mobile apps, personality traits, political beliefs and ideology, physical and mental health, demographic characteristics, and socioeconomic characteristics. There were 36 questions and the survey website was designed using Qualtrics (Qualtrics XM). We obtained research ethics approval for this project from the institutional review board of the Institute of Social Science, the University of Tokyo.

Measures

Uptake of Contact Tracing App

The app used a decentralized data privacy approach, which was among the 3 main design types (centralized, decentralized, and hybrid) for contact tracing apps worldwide [16,17]. Using Bluetooth sensors inside mobile phones, COCOA detects and records the app ID of other users who remain within 1 meter for more than 15 minutes [18]. The contact information is encrypted to maintain anonymity, and it is stored for 14 days only in the user's mobile phone before being automatically destroyed. This design secures users' privacy while tracing infection chains. In the event that a user is confirmed to be infected with COVID-19 and they voluntarily report it via the app, other users with whom they have been in close contact in the preceding 14 days receive a warning message. Individuals who receive a warning message can receive reverse-transcription polymerase chain reaction tests for free. Similar to apps in other countries, COCOA requires population adoption rates as high as 60% to contain the virus effectively. However, it is difficult to achieve this level through voluntary, individual compliance alone. As of December 28, 2020, the adoption rate was only 17.6% (22.5 million downloads) [19].

In this study, our dependent variable was a binary indicator (equal to 1 if the respondent had ever downloaded COCOA since its release in June 2020 and equal to 0 otherwise).

Prosociality

Individuals' motives for prosocial behavior have long been debated in various disciplines, including economics, sociology, and psychology. While different researchers categorize such motives in different manners, we draw on economics and psychology and classify them into intrinsic motivation, extrinsic incentive, and relationship with their community [20-22].

First, the literature on intrinsic motivation shows that individuals' prosocial behavior, such as volunteering, is attributed to their prosocial personality or preference, such as altruism, fairness, guilt, and empathy [23,24]. This is consistent with research from personality psychology on the agreeable personality trait, which includes facets of altruism and empathy [25,26].

Second, regarding the extrinsic incentive, people whose behavior deviates from the social norms of their community may experience nonmonetary punishment from others, such as disapproval, stigma, and negative social image [27]. Therefore, those who care about such punishments have incentives to behave prosocially. One may be concerned whether this motive is effective during the COVID-19 pandemic, given that research in criminology notes the aggravation of antisocial behavior in socially disorganized communities [28]. Although increases in crime in disaster-affected areas are a common problem worldwide [29,30], some have suggested the critical role of social norms and social images in Japanese disaster-affected communities [31]. Furthermore, social sanctions against antisocial behavior may be even stronger during the COVID-19 pandemic because of increased infection risk to community members. Hence, this motive may remain important in encouraging the uptake of contact tracing apps.

A third factor that motivates prosocial behavior is individuals' relationships with community members. Experimental studies [32-34] show that participants are more likely to be altruistic and cooperative when they play experimental games with in-group members, such as those sharing the same ethnicity and neighborhood. Empirical studies [35] also show that people are more likely to contribute to their community when its members are homogeneous in terms of ethnic and religious backgrounds, suggesting the importance of attachment to and identification with the community in encouraging prosocial behavior. These arguments are also in line with research on place identity [36], community attachment [37], and sense of community [38]. While there are some distinctions, many agree on the importance of emotional attachment to the community in motivating prosocial behavior [39,40].

To capture these motives for prosociality, we used 6 items. Items 1 and 2 (Table 1) have been proposed and validated [41] to elicit agreeableness in the Big 5 personality traits. These items were measured on a 7-point Likert scale, and a lower score on item 1 and a higher score on item 2 indicated higher agreeableness. Items 3 and 4 captured respondents' sensitivity to social norms. These were drawn from the World Values Survey [42] but were modified for ease of visibility on the web-based survey platform. These items were measured on a 5-point Likert scale. Items 5 and 6 measured respondents' attachment to their community. Item 5 demonstrated the highest factor loadings of 6 items in principal component analysis to capture place identity [43]. Other studies [44,45] also employ this item to measure place identity. Although this item measures individuals' place identity to their neighborhood, it may be the case that strength of identity varies with definition of place [46]; therefore, we added item 6 to capture identity as Japanese citizens. These items were measured on a 4-point Likert scale. Responses in which option 99 (ie, "do not want to answer") had been selected were dropped from the estimations. In line with methods used in earlier work [41], items 1 and 2 did not include option 99, but we allowed respondents to move to the next question without answering these questions.

Risk Perception

According to Protection Motivation Theory [47], risk perception describes how a person assesses a threat's probability and potential damage; it is determined based on perceived probability, perceived severity, fear, and the perceived reward for a maladaptive response. It has been suggested that variation in risk perception contributes to differences in behavioral responses to the COVID-19 pandemic in Japan [48].

We measured these characteristics with items 7 to 11 (Table 1). The first 4 questions measured perceived probability and severity of COVID-19 in terms of different domains, such as infection risk and job security. The fifth question is frequently used in the literature to measure individuals' willingness to take risks and draws from earlier work in the United States [49,50]. We used it as a proxy for fear. All items were measured with a 5-point Likert scale, where higher scores indicated higher risk perception.

Table 1. Description of items.

Items		Rating, mean (SD)
Prosociality		
1	I see myself as critical, quarrelsome.[a]	2.99 (1.41)
2	I see myself as sympathetic, warm.[a]	4.44 (1.27)
3	It is important to avoid doing anything people would say is wrong.[b]	2.96 (0.99)
4	It is important to behave properly.[b]	3.60 (0.94)
5	I am very attached to my neighborhood.[c]	2.80 (0.84)
6	I am proud of being a Japanese citizen.[c]	2.91 (0.81)
Risk perception		
7	I am concerned about the impact of COVID-19 on my infection risk.[b]	3.78 (1.02)
8	I am concerned about the impact of COVID-19 on serious symptoms.[b]	3.31 (1.10)
9	I am concerned about the impact of COVID-19 on my job.[b]	3.31 (1.20)
10	I am concerned about the impact of COVID-19 on my interpersonal relationships.[b]	3.21 (1.14)
11	Which of these sayings characterizes you better? (A) Nothing ventured, nothing gained (B) A wise man never courts danger[d]	3.60 (1.23)
Trust in government		
12	Do you trust the government?[c]	1.99 (0.79)
13	Do you evaluate the current prime minister positively?[b]	2.58 (1.20)
14	Do you evaluate the previous prime minister positively?[b]	2.29 (1.08)
15	Do you evaluate the current governor of home prefecture positively?[b]	2.94 (1.09)

[a]Response options were (1) disagree strongly, (2) disagree moderately, (3) disagree a little, (4) neither agree nor disagree, (5) agree a little, (6) agree moderately, or (7) agree strongly.

[b]Response options were (1) no, (2) weakly no, (3) neutral, (4) weakly yes, (5) yes, or (99) do not want to answer.

[c]Response options were (1) no, (2) weakly no, (3) weakly yes, (4) yes, or (99) do not want to answer.

[d]Response options were (1) B, (2) lean B, (3) neutral, (4) lean A, (5) A, or (99) do not want to answer.

Trust in Government

Studies [9,11] have noted that concerns about security and privacy are major obstacles to the adoption of contact tracing apps, given that such apps use GPS, Bluetooth, or other technologies that can reveal sensitive personal information. It is, therefore, unsurprising that individuals with low trust in the government (institutional trust) are less likely to use the apps [8,9]. Although the Japanese app prioritizes the protection of users' privacy from the government and corporations [51], there is anecdotal evidence that such concerns exist about the COCOA app [52].

Our survey included 4 questions to measure respondents' trust in government. Item 12 asked the extent to which respondents trust the Japanese government; responses were measured with a 4-point Likert scale. This is frequently used in the literature [7,53]. Items 13, 14, and 15 asked respondents to evaluate the performance of the current prime minister, previous prime minister, and the governor of their home prefecture, respectively, on a 5-point Likert scale.

Statistical Analysis

For data reduction, we performed 3 sets of exploratory factor analyses separately, using the items for prosociality, risk perception, and trust in government. Specifically, we used iterated principal factor extraction and promax rotation to obtain simple factor structures. The number of factors was determined based on the eigenvalue and the scree plot. To label the resulting factors, we used items with factor loadings above 0.4. Regression was used to calculate the factor score, and the factor score was standardized (mean 0, SD 1).

Understanding effective policy interventions to facilitate the uptake of contact tracing apps requires the analysis of both the determinants and consequences of these estimated factors. Therefore, we conducted ordinary least squares regression to examine socioeconomic and demographic predictors. Specifically, respondents' estimated factor scores were regressed on their age groups (20-29 years [baseline], 30-39 years, 40-49 years, 50-59 years, and 60-69 years), gender, whether the respondent completed university, whether the respondent engaged in a regular job (eg, self-employed, corporate executive, or full-time employee), marital status, whether they lived with

a child, whether they lived with a parent, and prefecture dummy variables. Japan has 47 prefectures, which are subnational units of government. The prefecture dummy variables were included to control for prefecture-level characteristics, such as the severity of infection spread. Standard errors were clustered at the prefecture level.

Subsequently, multivariate logistic regression was conducted to examine the association between the usage of COCOA and estimated factors for prosociality, risk perception, and trust in the government. The estimation model also included the control variables (respondents' age, gender, completed university, a regular job, marital status, cohabitation with a child, cohabitation with a parent, and prefecture dummies). Results from regression analyses were reported as odds ratio (OR) with 95% CI. In addition to the full-sample model, we conducted subsample estimations by respondent age groups (3 categories: 20-39 years, 40-59 years, and 60-69 years).

Given that the empirical results depend on the method used to quantify these factors, for robustness, we used 2 alternative approaches for the logistic regression models. First, for each factor, we chose the item with the highest factor loading and used the responses to these items as independent variables instead of factor scores. Second, using only the items whose factor loadings were higher than 0.4 in the factor analysis, we conducted principal component analysis with 1 component. Subsequently, we estimated the predicted scores and used them in the logistic regression. All analyses were performed in Stata (version 14; StataCorp LLC).

Results

Sample Characteristics

Among the 7084 respondents who completed the survey, we discarded the responses of those who finished the survey too quickly (less than 5 minutes) or too slowly (more than 30 minutes) to control for survey quality. Our final sample size was 5402, of which 791 respondents (14.6%) had used COCOA. Males, university graduates, and those with regular jobs were more likely to use COCOA (Table 2). Differences were not significant for age (P=.09), marital status (P=.39), or household structure (living with child: P=.15; living with parent: P=.19).

Table 2. Characteristics of the study sample grouped by COCOA usage.

Characteristic	All (N=5402)	Users (n=791)	Non-users (n=4611)	*P* value
Age (years), mean (SD)	45.94 (13.28)	46.68 (13.44)	45.81 (13.25)	.09
Gender, n (%)				.006
Female	2544 (47.1)	337 (42.6)	2207 (47.9)	
Male	2858 (52.9)	454 (57.4)	2404 (52.1)	
Education, n (%)				<.001
Completed university	2775 (51.4)	468 (59.2)	2307 (50.0)	
Did not complete university	2627 (48.6)	323 (40.8)	2304 (50.0)	
Employment, n (%)				<.001
Regular job	2725 (50.4)	454 (57.4)	2271 (49.3)	
Nonregular job	1130 (20.9)	152 (19.2)	978 (21.2)	
Not working	1494 (27.7)	178 (22.5)	1316 (28.5)	
Other	53 (1.0)	7 (0.9)	46 (1.0)	
Marital status, n (%)				.39
Married	3031 (56.1)	455 (57.5)	2576 (55.9)	
Unmarried	1988 (36.8)	280 (35.4)	1708 (37.1)	
Other	377 (7.0)	56 (7.1)	321 (7.0)	
Living with parent, n (%)				.19
Yes	1491 (27.6)	203 (25.7)	1288 (27.9)	
No	3911 (72.4)	588 (74.3)	3323 (72.1)	
Living with child, n (%)				.15
Yes	1798 (33.3)	281 (35.5)	1517 (32.9)	
No	3604 (66.7)	510 (64.5)	3094 (67.1)	

To evaluate the national representativeness of our respondents, we compared the characteristics of our respondents with those of smartphone owners in Japan from the Communications Usage Trend Survey [54], a nationally representative survey conducted by the government in 2019. This survey collected information about the usage of Information and Communication Technologies among Japanese citizens [54]. Since this study examines uptake decisions for a mobile app, those who do not

have access to internet or smartphone were not of interest. Our sample of respondents was representative of smartphone owners in Japan (Table S1 in Multimedia Appendix 1).

Factor Analysis

We began by conducting a Barlett test of sphericity, which found significant correlations between items ($P<.001$), suggesting the adequacy of using factor analysis. Regarding prosociality, although only the first factor demonstrated an eigenvalue greater than 1 (eigenvalue 1.52), we also retained the second factor (eigenvalue 0.75) based on the scree plot (Table 3). These factors explained 63.1% and 31.0% of the variance in the data. After promax rotation, items 1 and 2 demonstrated high factor loadings in the first factor, which we labeled *Agreeableness*. The second factor was characterized by high factor loadings of items 5 and 6; therefore, it was labeled *Attachment to the community*.

We extracted 2 factors for risk perception (Table 4). The first factor (eigenvalue 1.80) demonstrated high factor loadings for items 7 and 8; therefore, it was labeled *Concern about health risk*. Likewise, the second factor (eigenvalue 0.47), which demonstrated high factor loadings for items 9 and 10, was labeled *Concern about social risk*. These factors accounted for 74.1% and 19.3% of the variance.

Finally, we extracted 1 factor related to trust in government (eigenvalue 1.89). Given the high factor loadings of items 12, 13, and 14 and the low factor loading of item 15, it was labeled *Trust in national government*. This accounted for 98.6% of total variance (Table 5).

Table 3. Prosociality factor loadings.

Prosociality items	Agreeableness	Attachment to the community
I see myself as critical, quarrelsome.	–0.5488	–0.0278
I see myself as sympathetic, warm.	0.5365	0.0007
I am very attached to my neighborhood.	–0.0046	0.7129
I am proud of being a Japanese citizen.	0.0155	0.7043
It is important to avoid doing anything people would say is wrong.	0.0167	–0.0052
It is important to behave properly.	0.2500	0.0071

Table 4. Risk perception factor loadings.

Risk perception items	Concern about health risk	Concern about social risk
I am concerned about the impact of COVID-19 on serious symptoms.	0.8329	–0.0113
I am concerned about the impact of COVID-19 on my infection risk.	0.6532	0.0825
I am concerned about the impact of COVID-19 on my interpersonal relationships.	0.0633	0.6817
I am concerned about the impact of COVID-19 on my job.	–0.0139	0.6371
Which of these sayings characterizes you better? (A) Nothing ventured, nothing gained (B) A wise man never courts danger	–0.0251	–0.0259

Table 5. Trust in government factor loadings.

Trust in government items	Trust in national government
Do you evaluate the current prime minister positively?	0.7853
Do you evaluate the previous prime minister positively?	0.7695
Do you trust the government?	0.4557
Do you evaluate the current governor of home prefecture positively?	0.2325

Predictors of Factor Scores

Factor score variables were standardized. Older and married respondents had higher agreeableness and attachment to the community (Table 6). In addition, female and university-educated respondents demonstrated higher agreeableness scores, and respondents cohabiting with a parent or a child exhibited higher attachment to the community. The factors for risk perception were higher for female respondents and respondents cohabiting with a parent. Concern about health risk increased with age and was higher for married persons, while concern about social risk was higher for respondents with higher educational attainment, with a regular job, and who cohabited with a child. Finally, trust in national government was negatively associated with age and positively associated with cohabiting with a child.

Table 6. Predictors of factor scores.

Variable	Ordinary least square coefficients (95% CI)				
	Agreeableness	Attachment to the community	Concern about health risk	Concern about social risk	Trust in national government
Aged 30 to 39 years	–0.132** (–0.210 to –0.055)	–0.096 (–0.217 to 0.024)	0.04 (–0.052 to 0.126)	–0.087 (–0.182 to 0.007)	–0.194*** (–0.264 to –0.123)
Aged 40 to 49 years	–0.062 (–0.137 to 0.012)	–0.005 (–0.126 to 0.116)	0.130** (0.052 to 0.207)	–0.009 (–0.108 to 0.091)	–0.300*** (–0.378 to –0.221)
Aged 50 to 59 years	0.103** (0.027 to 0.179)	0.190*** (0.088 to 0.293)	0.258*** (0.172 to 0.344)	–0.017 (–0.111 to 0.076)	–0.312*** (–0.398 to –0.226)
Aged 60 to 69 years	0.250*** (0.156 to 0.345)	0.446*** (0.327 to 0.564)	0.435*** (0.355 to 0.515)	–0.074 (–0.184 to 0.036)	–0.401*** (–0.499 to –0.302)
Female	0.177*** (0.129 to 0.226)	0.07 (–0.008 to 0.155)	0.215*** (0.161 to 0.270)	0.216*** (0.161 to 0.272)	0.01 (–0.066 to 0.080)
Completed university	0.096** (0.034 to 0.158)	0.00 (–0.073 to 0.081)	0.05 (–0.001 to 0.099)	0.081** (0.029 to 0.132)	–0.001 (–0.059 to 0.057)
Regular job	0.02 (–0.040 to 0.070)	0.05 (–0.012 to 0.112)	0.01 (–0.046 to 0.058)	0.320*** (0.264 to 0.376)	0.06 (–0.013 to 0.125)
Married	0.123** (0.048 to 0.198)	0.126*** (0.062 to 0.190)	0.092* (0.020 to 0.164)	0.05 (–0.042 to 0.131)	0.02 (–0.044 to 0.078)
Live with a parent	0.03 (–0.035 to 0.092)	0.121*** (0.065 to 0.176)	0.127** (0.052 to 0.202)	0.096** (0.028 to 0.164)	0.02 (–0.051 to 0.091)
Live with a child	0.06 (–0.010 to 0.132)	0.143*** (0.071 to 0.214)	0.04 (–0.042 to 0.124)	0.09 (–0.002 to 0.191)	0.095** (0.032 to 0.159)
Prefecture fixed effects	Yes	Yes	Yes	Yes	Yes

*P<.05.
**P<.01.
***P<.001.

Association With the Uptake of COCOA

Full-sample results (Table 7) for Model 1 show that all factors significantly increased the odds of using COCOA (agreeableness: OR 1.15, 95% CI 1.07-1.23; attachment to the community: OR 1.20, 95% CI 1.07-1.35; concern about health risk: OR 1.25, 95% CI 1.12-1.41; concern about social risk: OR 1.14, 95% CI 1.04-1.24; trust in national government: OR 1.08, 95% CI 1.00-1.17). OR magnitudes did not differ significantly across factors (χ^2=6.30, P=.18). Respondents' socioeconomic status characteristics, such as education and having a regular job, were positively correlated with app usage. University graduates were 1.33 times more likely to install COCOA than high school graduates were, and having a regular job increased odds by 1.25. Other demographic characteristics (gender: P=.12; age: P=.25; living with a parent: P=.41, living with a child: P=.41) were not correlated with app uptake. Subsample estimations for Models 2 to 4 uncovered heterogeneity in determinants across the 3 major age groups. Among respondents aged 20 to 39 years (Model 2), attachment to the community was a major determinant of uptake (OR 1.28, 95% CI 1.11-1.48). Among respondents aged 40 to 59 years, the uptake of COCOA was attributed to respondents' high agreeableness (OR 1.18, 95% CI 1.02-1.35), concern about social risk (OR 1.17, 95% CI 1.02-1.35), and trust in the national government (OR 1.16, 95% CI 1.05-1.28). For respondents aged over 60, uptake was attributed to higher concern about health risks (OR 1.49, 95% CI 1.24-1.80).

The correlates of COCOA uptake using responses to items 1, 5, 8, 10, and 13, which had the highest factor loadings for each of the factors, demonstrate robustness (Table S2 in Multimedia Appendix 1), and the association between COCOA uptake and factor scores computed with principal component analysis did not change qualitatively (Table S3 in Multimedia Appendix 1).

Table 7. Association with the uptake of the COVID-19 Contact Confirming App.

Variable	Odds ratio (95% CI)			
	Model 1: All (n=5398)	Model 2: Age 20-39 years (n=1765)	Model 3: Age 40-59 years (n=2488)	Model 4: Age 60-69 years (n=982)
Agreeableness	1.15*** (1.07 to 1.23)	1.13 (0.99 to 1.30)	1.18* (1.02 to 1.35)	1.15 (0.97 to 1.36)
Attachment to the community	1.20** (1.07 to 1.35)	1.28*** (1.11 to 1.48)	1.19 (0.94 to 1.50)	1.03 (0.81 to 1.30)
Concern about health risk	1.25*** (1.12 to 1.41)	1.15 (0.94 to 1.42)	1.20 (0.97 to 1.48)	1.49*** (1.24 to 1.80)
Concern about social risk	1.14** (1.04 to 1.24)	1.18 (0.96 to 1.45)	1.17* (1.02 to 1.35)	1.05 (0.85 to 1.29)
Trust in national government	1.08* (1.00 to 1.17)	0.98 (0.85 to 1.14)	1.16** (1.05 to 1.28)	1.08 (0.88 to 1.31)
Age	1.00 (1.00 to 1.01)	0.99 (0.95 to 1.02)	1.01 (0.98 to 1.03)	1.00 (0.93 to 1.08)
Female	0.88 (0.76 to 1.03)	0.93 (0.70 to 1.24)	0.92 (0.77 to 1.10)	0.74 (0.45 to 1.22)
Completed university	1.33** (1.12 to 1.57)	1.77*** (1.30 to 2.42)	1.19 (0.95 to 1.49)	1.07 (0.65 to 1.76)
Regular job	1.25** (1.07 to 1.47)	1.11 (0.77 to 1.60)	1.36* (1.01 to 1.83)	1.32 (0.87 to 2.00)
Married	0.89 (0.74 to 1.08)	0.78 (0.50 to 1.22)	0.97 (0.75 to 1.26)	0.94 (0.67 to 1.31)
Live with a parent	0.93 (0.79 to 1.10)	0.96 (0.70 to 1.30)	0.95 (0.71 to 1.27)	0.86 (0.34 to 2.16)
Live with a child	1.07 (0.91 to 1.26)	1.13 (0.68 to 1.88)	1.10 (0.88 to 1.36)	1.10 (0.76 to 1.60)
Prefecture fixed effects	Yes	Yes	Yes	Yes
Hosmer-Lemeshow *P* value	.997	.450	.188	.161

*$P<.05$.
**$P<.01$.
***$P<.001$.

Discussion

Principal Results

Using a unique survey in Japan, we found that individuals' uptake of COVID-19 contact tracing apps is determined by their agreeableness, attachment to the community, concern about health risks, concern about social risks, and trust in the national government; however, key determinants differ across generations. For cohorts aged between 20 and 39 years, attachment to the community plays a pivotal role, while concerns about their health, the social impact of COVID-19, and trust in the national government are less relevant. For those aged between 40 and 59 years, an agreeable personality, concern about the social impact of COVID-19, and trust in the national government facilitate uptake. Finally, adults over 60 years of age, having greater concern about the health impact of COVID-19, were more likely to download the app.

Providing rigorous evidence to explain the causes of heterogeneous patterns across generations is a challenge. That said, we speculate that downloading the contact tracing app may offer fewer benefits for younger age groups, who are less likely to become severely ill from COVID-19. Hence, they may see it primarily as prosocial behavior, making strong attachment to the community at local and national levels essential for uptake. Furthermore, we found that trust in government did not influence uptake decisions for younger age groups, likely because they already use many mobile apps, including those for web-based games, shopping, and social media, and thus may be less concerned about online privacy and security. In contrast, health care is the largest issue among older adults, and those who were concerned about their health risk used the app, regardless of their trust in government or prosociality. Finally, middle-age respondents were more likely than those in other age groups to live in a large household with children and parents; therefore, they may have been more concerned about security and privacy issues. Hence, their uptake decision relied mostly on whether they find the government trustworthy or not. If they did not trust the government, even individuals who were prosocial and concerned about health risks were less likely to install COCOA.

Limitations

A potential limitation of this study was sample selection. We used a web-based survey because the situation created by the spread of COVID-19 made it difficult to conduct either paper-and-pencil postal surveys or in-person surveys in a timely manner. As a result, those with poor internet literacy were excluded from our sample. However, we believe that this issue is unlikely to be severe. Since this study examines the uptake decision with respect to a mobile app, those who do not have access to the internet or smartphones are less relevant to our analysis. Furthermore, the characteristics of our respondents were comparable with those of smartphone owners in Japan.

Comparison With Prior Work

This study makes 4 contributions to the literature. First, the importance of prosociality and community in controlling the COVID-19 pandemic has been frequently discussed, and previous studies [55,56] have demonstrated the association of prosociality and community with individuals' social-distancing

behavior and mental health. In general, the literature on public health and disaster research also supports the roles of community in encouraging protective behavior [14,57]. However, the association between prosociality and uptake of COVID-19 contact tracing apps is largely unexplored, with the exception of 2 studies that have examined the role of prosocial personality traits and attitudes [7,58]; attachment to the community is less well understood. This is problematic because, unlike personality traits, feelings of attachment may easily change over time in response to changes in one's living conditions [59]. Our findings suggest that, if social-distancing requirements during the pandemic weaken community attachment, then this could have a negative effect on uptake decisions, especially among young people.

Second, this study is the first to examine differences in app uptake across generations. Differences across age groups suggest that conducting an empirical analysis without taking generational heterogeneity into consideration (as previous studies have neglected to take into consideration) leads to misunderstandings about individuals' behavioral responses to the COVID-19 pandemic. Risk and symptomatic severity of infections vary across age groups. Specifically, young generations account for a large proportion of confirmed cases in Japan, and it is relevant for policymakers to contain the spread of infection among young generations even though they are less likely to become severely ill.

Third, previous studies [9,11,53] on contact tracing apps mainly used survey data collected before the release of the apps and examined the willingness to install a hypothetical app. While their arguments were insightful, the actual adoption rates were remarkably lower than what had been predicted. This suggests the importance of further research to analyze actual uptake decisions, as performed in our study and in a few others [7,8,60].

Fourth, this study contributed to the literature on disaster resilience. Existing scholarship emphasizes the importance of strong communities and institutions, along with the development of physical infrastructure [61,62]. However, whether these factors play complementary or substitute roles remains unsubstantiated. This study provides rigorous evidence that attachment to one's community boosts the effectiveness of contact tracing technology.

Conclusion

Given these arguments, policymakers should implement and advance different interventions for each generation to increase the adoption rate of contact tracing apps. These strategies are relevant, not only to the COVID-19 pandemic, but also, to possible future pandemics in which decentralized contract-tracing may be relevant to the mitigation of human, social, and economic suffering. Specifically, older adults demonstrate higher concerns about health risks than younger individuals; such concerns are the primary motivation for uptake by older adults. Therefore, a promising approach is to inform them about the health benefit from the apps, such as receiving medical treatment sooner. For middle-age persons, it is important to mitigate their concerns about security and privacy issues. Finally, uptake by young persons is determined by their attachment to the community; however, interventions to inform them that the app prevents users from spreading the infection may not be effective, because on average, young adults do not feel as attached to their community as older adults. Instead, it is important to maintain and raise their feeling of community attachment at the local and national levels. This may be challenging because social-distancing requirements during the pandemic have reduced face-to-face social interactions among community members. However, the use of social media and other web-based network tools may compensate for the lack of such opportunities. A study [63] of American university students showed that communication through social media (Facebook) helped young people maintain relationships with those who were physically at a distance, such as high school friends from their hometowns. Some Japanese municipalities have introduced web-based events to facilitate social interactions among the young generation during the COVID-19 pandemic, such as coming-of-age ceremonies, childcare workshops for young parents, and festivals or activities for families with young children [64]. That said, these approaches also have drawbacks. First, online or social media communities and online meetings often include only young users, whose health risk is low. Social interactions among such people may not lead to stronger motivations to engage in prosocial behavior for senior persons. Second, it is not evident how long prosociality developed through web-based tools persists.

While these implications are grounded in evidence from Japan, we expect them to be pertinent to other countries. The relevance of risk perception and trust in government to uptake decisions for contact tracing apps have been widely recognized in many countries [7-9,11]. In addition, our argument about the role of prosociality is applicable to any country where app uptake depends on voluntary, individual decisions alone. Nonetheless, we should be cautious about the generalizability of implementing different interventions across generations, because there is no comparable evidence from other countries on generational differences in uptake decisions. The key determinants of uptake among young and older generations may depend on the demographic, cultural, and socioeconomic characteristics of each country. Additional studies in other countries are required to establish which combination of policy interventions is most effective for each generation.

Acknowledgments

This manuscript was prepared with support from the Institute of Social Science (University of Tokyo) and its institute-wide joint research project (Methodology of Social Sciences: How to Measure Phenomena and Values). This work was supported by Japan Society for the Promotion of Science Kakenhi grants (20K01689, 18K01501, and 17H02478) and the Institute of Social Science Group Joint Research Project grant (Social Sciences of Digitalization).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[DOCX File , 29 KB - mhealth_v9i8e29923_app1.docx]

References

1. Atkeson A. What will be the economic impact of COVID-19 in the US? rough estimates of disease scenarios. National Bureau of Economic Research. 2020 Mar. URL: https://www.nber.org/system/files/working_papers/w26867/w26867.pdf [accessed 2021-08-15]
2. Sumner A, Hoy C, Ortiz-Juarez E. Estimates of the Impact of COVID-19 on Global Poverty. United Nations University World Institute for Development Economics Research. 2020. URL: https://www.wider.unu.edu/sites/default/files/Publications/Working-paper/PDF/wp2020-43.pdf [accessed 2021-08-15]
3. Budd J, Miller BS, Manning EM, Lampos V, Zhuang M, Edelstein M, et al. Digital technologies in the public-health response to COVID-19. Nat Med 2020 Aug 07;26(8):1183-1192. [doi: 10.1038/s41591-020-1011-4] [Medline: 32770165]
4. Johnson B. The COVID tracing tracker: what's happening in coronavirus apps around the world. MIT Technology Review. 2020 Dec 16. URL: https://www.technologyreview.com/2020/12/16/1014878/covid-tracing-tracker/ [accessed 2021-05-20]
5. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 2020 May 08;368(6491):eabb6936 [FREE Full text] [doi: 10.1126/science.abb6936] [Medline: 32234805]
6. Hinch R, Probert W, Nurtay A, Kendall M, Wymant C, Hall M, et al. Effective configurations of a digital contact tracing app: a report to NHSX. Semantic Scholar. 2020. URL: https://cdn.theconversation.com/static_files/files/1009/Report_-_Effective_App_Configurations.pdf?1587531217 [accessed 2021-05-20]
7. Munzert S, Selb P, Gohdes A, Stoetzer LF, Lowe W. Tracking and promoting the usage of a COVID-19 contact tracing app. Nat Hum Behav 2021 Feb 21;5(2):247-255. [doi: 10.1038/s41562-020-01044-x] [Medline: 33479505]
8. von Wyl V, Höglinger M, Sieber C, Kaufmann M, Moser A, Serra-Burriel M, et al. Drivers of acceptance of COVID-19 proximity tracing apps in Switzerland: panel survey analysis. JMIR Public Health Surveill 2021 Jan 06;7(1):e25701 [FREE Full text] [doi: 10.2196/25701] [Medline: 33326411]
9. Altmann S, Milsom L, Zillessen H, Blasone R, Gerdon F, Bach R, et al. Acceptability of app-based contact tracing for COVID-19: cross-country survey study. JMIR Mhealth Uhealth 2020 Aug 28;8(8):e19857 [FREE Full text] [doi: 10.2196/19857] [Medline: 32759102]
10. Walrave M, Waeterloos C, Ponnet K. Adoption of a contact tracing app for containing COVID-19: a health belief model approach. JMIR Public Health Surveill 2020 Sep 01;6(3):e20572 [FREE Full text] [doi: 10.2196/20572] [Medline: 32755882]
11. Jonker M, de Bekker-Grob E, Veldwijk J, Goossens L, Bour S, Rutten-Van Mölken M. COVID-19 contact tracing apps: predicted uptake in the Netherlands based on a discrete choice experiment. JMIR Mhealth Uhealth 2020 Oct 9;8(10):e20741. [doi: 10.2196/20741]
12. Piliavin A, Dovidio F, Gaertner L, Clark D. Emergency Intervention. New York: Academic Press; 1981.
13. Mazza M, Marano G, Lai C, Janiri L, Sani G. Danger in danger: interpersonal violence during COVID-19 quarantine. Psychiatry Res 2020 Apr 30;289:113046. [doi: 10.1016/j.psychres.2020.113046] [Medline: 33242810]
14. Aldrich D. Building Resilience: Social Capital in Postdisaster Recovery. Chicago, IL: University of Chicago Press; 2012.
15. COVID-19. Toyo Keizai Online. URL: https://toyokeizai.net/sp/visual/tko/covid19/ [accessed 2021-04-07]
16. Ahmed N, Michelin RA, Xue W, Ruj S, Malaney R, Kanhere SS, et al. A survey of COVID-19 contact tracing apps. IEEE Access 2020;8:134577-134601. [doi: 10.1109/access.2020.3010226]
17. Nakamoto I, Jiang M, Zhang J, Zhuang W, Guo Y, Jin M, et al. Evaluation of the design and implementation of a peer-to-peer COVID-19 contact tracing mobile app (COCOA) in Japan. JMIR Mhealth Uhealth 2020 Dec 01;8(12):e22098 [FREE Full text] [doi: 10.2196/22098] [Medline: 33170801]
18. Contact confirmation application privacy policy. Ministry of Health, Labor and Welfare, Japan. 2020 Dec 15. URL: https://www.mhlw.go.jp/cocoa/privacypolicy_english.html [accessed 2021-05-20]
19. COCOA official home page. Ministry of Health, Labor and Welfare, Japan. URL: https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/cocoa_00138.html [accessed 2021-04-19]
20. Prendergast C. Intrinsic motivation and incentives. American Economic Review 2008 Apr 01;98(2):201-205. [doi: 10.1257/aer.98.2.201]
21. Cerasoli CP, Nicklin JM, Ford MT. Intrinsic motivation and extrinsic incentives jointly predict performance: a 40-year meta-analysis. Psychol Bull 2014 Jul;140(4):980-1008. [doi: 10.1037/a0035661] [Medline: 24491020]
22. Deci EL, Koestner R, Ryan RM. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. Psychol Bull 1999 Nov;125(6):627-68; discussion 692. [doi: 10.1037/0033-2909.125.6.627] [Medline: 10589297]

23. Carpenter J, Myers CK. Why volunteer? evidence on the role of altruism, image, and incentives. J Public Econ 2010 Dec;94(11-12):911-920. [doi: 10.1016/j.jpubeco.2010.07.007]
24. Shoji M. Guilt and prosocial behavior: lab-in-the-field evidence from Bangladesh. Econ Dev Cult Change 2021 Feb 17:1 (forthcoming). [doi: 10.1086/713879]
25. Graziano G, Eisenberg N. Agreeableness: a dimension of personality. In: Hogan R, Johnson J, Briggs S, editors. Handbook of Personality Psychology. Cambridge, MA: Academic Press; 1997:795-824.
26. Ashton MC, Paunonen SV, Helmes E, Jackson DN. Kin altruism, reciprocal altruism, and the big five personality factors. Evol Hum Behav 1998 Jul;19(4):243-255. [doi: 10.1016/s1090-5138(98)00009-9]
27. Bursztyn L, Jensen R. Social image and economic behavior in the field: identifying, understanding, and shaping social pressure. Annu Rev Econ 2017 Aug 02;9(1):131-153. [doi: 10.1146/annurev-economics-063016-103625]
28. Shaw CR, McKay HD. Juvenile Delinquency and Urban Areas. Chicago, Illinois: University of Chicago Press; Jan 1943:681.
29. Frailing K, Wood-Harper D. Crime and disaster in historical perspective. In: Crime and Criminal Justice in Disaster. Durham NC: Carolina Academic Press; 2010.
30. Shoji M. Religious fractionalisation and crimes in disaster-affected communities: survey evidence from Bangladesh. J Dev Stud 2017 Nov 06;54(10):1891-1911. [doi: 10.1080/00220388.2017.1393521]
31. Kondo A, Shoji M. Peer effects in employment status: evidence from housing lotteries. J Urban Econ 2019 Sep;113:103195. [doi: 10.1016/j.jue.2019.103195]
32. Ben-Ner A, McCall BP, Stephane M, Wang H. Identity and in-group/out-group differentiation in work and giving behaviors: Experimental evidence. J Econ Behav Organ 2009 Oct;72(1):153-170. [doi: 10.1016/j.jebo.2009.05.007]
33. Fershtman C, Gneezy U. Discrimination in a Segmented Society: An Experimental Approach. Q J Econ 2001 Feb 01;116(1):351-377. [doi: 10.1162/003355301556338]
34. Etang A, Fielding D, Knowles S. Does trust extend beyond the village? experimental trust and social distance in Cameroon. Exp Econ 2010 Sep 16;14(1):15-35. [doi: 10.1007/s10683-010-9255-3]
35. Alesina A, La Ferrara E. Participation in Heterogeneous Communities*. Q J Econ 2000 Aug;115(3):847-904. [doi: 10.1162/003355300554935]
36. Proshansky HM, Fabian AK, Kaminoff R. Place-identity: physical world socialization of the self. J Environ Psychol 1983 Mar;3(1):57-83. [doi: 10.1016/s0272-4944(83)80021-8]
37. Kasarda JD, Janowitz M. Community attachment in mass society. Am Sociol Rev 1974 Jun;39(3):328. [doi: 10.2307/2094293]
38. McMillan DW, Chavis DM. Sense of community: a definition and theory. J Community Psychol 1986 Jan;14(1):6-23. [doi: 10.1002/1520-6629(198601)14:1<6::aid-jcop2290140103>3.0.co;2-i]
39. Brown G, Raymond C. The relationship between place attachment and landscape values: toward mapping place attachment. Appl Geogr 2007 Apr;27(2):89-111. [doi: 10.1016/j.apgeog.2006.11.002]
40. Mesch GS, Manor O. Social ties, environmental perception, and local attachment. Environ Behav 2016 Jul 26;30(4):504-519. [doi: 10.1177/001391659803000405]
41. Gosling SD, Rentfrow PJ, Swann WB. A very brief measure of the big-five personality domains. J Res Pers 2003 Dec;37(6):504-528. [doi: 10.1016/s0092-6566(03)00046-1]
42. Inglehart R, Haerpfer C, Moreno A, Welzel C, Kizilova K, Diez-Medrano J, et al. Round six - country-pooled datafile version. World Values Survey. JD Systems Institute: Madrid; 2014. URL: https://www.worldvaluessurvey.org/WVSDocumentationWV6.jsp [accessed 2021-08-15]
43. Raymond CM, Brown G, Weber D. The measurement of place attachment: personal, community, and environmental connections. J Environ Psychol 2010 Dec;30(4):422-434. [doi: 10.1016/j.jenvp.2010.08.002]
44. Pitas NAD, Mowen AJ, Powers SL. Person-place relationships, social capital, and health outcomes at a nonprofit community wellness center. J Leis Res 2020 Jun 16;52(2):247-264. [doi: 10.1080/00222216.2020.1776652]
45. Williams DR, Patterson ME, Roggenbuck JW, Watson AE. Beyond the commodity metaphor: examining emotional and symbolic attachment to place. Leis Sci 1992 Jan;14(1):29-46. [doi: 10.1080/01490409209513155]
46. Gustafson P. More cosmopolitan, no less local: the orientations of international travellers. Eur Soc 2009 Feb;11(1):25-47. [doi: 10.1080/14616690802209689]
47. Rogers W, Prentice-Dunn S. Protection motivation theory. In: Handbook of Health Behavior Research 1: Personal and Social Determinants. New York, NY: Plenum Press; 1997:113-132.
48. Shoji M, Cato S, Iida T, Ishida K, Ito A, McElwain K. COVID-19 and social distancing in the absence of legal enforcement: survey evidence from Japan. University Library LMU Munich. 2020. URL: https://mpra.ub.uni-muenchen.de/101968/ [accessed 2021-08-15]
49. Ikeda S, Kato H, Ohtake F, Tsutsui Y. Behavioral Economics of Preferences, Choices, and Happiness. New York NY: Springer; 2016.
50. Iida T. Yūkensha no Risuku Taido to Tōhyōkōdō. Tokyo: Bokutakusha; 2016.
51. Nakamoto I, Jiang M, Zhang J, Zhuang W, Guo Y, Jin M, et al. Evaluation of the design and implementation of a peer-to-peer COVID-19 contact tracing mobile app (COCOA) in Japan. JMIR Mhealth Uhealth 2020 Dec 01;8(12):e22098 [FREE Full text] [doi: 10.2196/22098] [Medline: 33170801]

52. Sesshoku Kakunin Apuri Hukyuu Habamu Kabe wo Dou Torinozokuka. August. In: Yomiuri Shimbun. Tokyo: Yomiuri Shimbun; Aug 3, 2020.
53. Guillon M, Kergall P. Attitudes and opinions on quarantine and support for a contact-tracing application in France during the COVID-19 outbreak. Public Health 2020 Nov;188:21-31 [FREE Full text] [doi: 10.1016/j.puhe.2020.08.026] [Medline: 33059232]
54. Communications usage trend survey. Ministry of Internal Affairs and Communications. 2019. URL: https://www.soumu.go.jp/johotsusintokei/statistics/pdf/HR201900_001.pdf [accessed 2021-06-03]
55. Cato S, Iida T, Ishida K, Ito A, McElwain K, Shoji M. Social distancing as a public good under the COVID-19 pandemic. Public Health 2020 Nov;188:51-53 [FREE Full text] [doi: 10.1016/j.puhe.2020.08.005] [Medline: 33120232]
56. Meagher BR, Cheadle AD. Distant from others, but close to home: the relationship between home attachment and mental health during COVID-19. J Environ Psychol 2020 Dec;72:101516. [doi: 10.1016/j.jenvp.2020.101516]
57. Kawachi I, Subramanian V, Kim D. , ,. In: Social Capital and Health. New York, NY: Springer; 2008.
58. Li T, Cobb C, Yang J, Baviskar S, Agarwal Y, Li B, et al. What makes people install a COVID-19 contact-tracing app? understanding the influence of app design and individual difference on contact-tracing app adoption intention. Pervasive Mob Comput 2021 Aug;75:101439. [doi: 10.1016/j.pmcj.2021.101439]
59. von Wirth T, Grêt-Regamey A, Moser C, Stauffacher M. Exploring the influence of perceived urban change on residents' place attachment. J Environ Psychol 2016 Jun;46:67-82. [doi: 10.1016/j.jenvp.2016.03.001]
60. Saw YE, Tan EY, Liu JS, Liu JC. Predicting public uptake of digital contact tracing during the COVID-19 pandemic: results from a nationwide survey in Singapore. J Med Internet Res 2021 Feb 03;23(2):e24730 [FREE Full text] [doi: 10.2196/24730] [Medline: 33465034]
61. Ainuddin S, Routray JK. Community resilience framework for an earthquake prone area in Baluchistan. Int J Disaster Risk Reduct 2012 Dec;2:25-36. [doi: 10.1016/j.ijdrr.2012.07.003]
62. Cutter SL, Barnes L, Berry M, Burton C, Evans E, Tate E, et al. A place-based model for understanding community resilience to natural disasters. Glob Environ Change 2008 Oct;18(4):598-606. [doi: 10.1016/j.gloenvcha.2008.07.013]
63. Ellison B, Steinfield C, Lampe C. The benefits of Facebook “friends:” social capital and college students’ use of online social network sites. J Comput Mediat Commun 2007;12(4):1143-1168. [doi: 10.1111/j.1083-6101.2007.00367.x]
64. Atarashii Chiiki Katsudou Sutairu Jireishuu. Kyoto City Official Website. URL: https://www.city.kyoto.lg.jp/bunshi/page/00002683html [accessed 2021-06-03]

Abbreviations

COCOA: COVID-19 Contact Confirming App
OR: odds ratio

Edited by L Buis; submitted 26.04.21; peer-reviewed by N Lunt, X Cheng, E Aono; comments to author 19.05.21; revised version received 07.06.21; accepted 23.07.21; published 19.08.21.

Please cite as:
Shoji M, Ito A, Cato S, Iida T, Ishida K, Katsumata H, McElwain KM
Prosociality and the Uptake of COVID-19 Contact Tracing Apps: Survey Analysis of Intergenerational Differences in Japan
JMIR Mhealth Uhealth 2021;9(8):e29923
URL: https://mhealth.jmir.org/2021/8/e29923
doi:10.2196/29923
PMID:34313601

Original Paper

Comparison of the Validity and Generalizability of Machine Learning Algorithms for the Prediction of Energy Expenditure: Validation Study

Ruairi O'Driscoll[1], MSc; Jake Turicchi[1], BSc; Mark Hopkins[2], PhD; Cristiana Duarte[1], PhD; Graham W Horgan[3], PhD; Graham Finlayson[1], PhD; R James Stubbs[1], PhD

[1]Appetite Control and Energy Balance Group, School of Psychology, University of Leeds, Leeds, United Kingdom

[2]School of Food Science and Nutrition, Faculty of Environment, University of Leeds, Leeds, United Kingdom

[3]Biomathematics & Statistics Scotland, Aberdeen, United Kingdom

Corresponding Author:
Ruairi O'Driscoll, MSc
Appetite Control and Energy Balance Group
School of Psychology
University of Leeds
Woodhouse
Leeds
United Kingdom
Phone: 44 113 343 2846
Email: psrod@leeds.ac.uk

Abstract

Background: Accurate solutions for the estimation of physical activity and energy expenditure at scale are needed for a range of medical and health research fields. Machine learning techniques show promise in research-grade accelerometers, and some evidence indicates that these techniques can be applied to more scalable commercial devices.

Objective: This study aims to test the validity and out-of-sample generalizability of algorithms for the prediction of energy expenditure in several wearables (ie, Fitbit Charge 2, ActiGraph GT3-x, SenseWear Armband Mini, and Polar H7) using two laboratory data sets comprising different activities.

Methods: Two laboratory studies (study 1: n=59, age 44.4 years, weight 75.7 kg; study 2: n=30, age=31.9 years, weight=70.6 kg), in which adult participants performed a sequential lab-based activity protocol consisting of resting, household, ambulatory, and nonambulatory tasks, were combined in this study. In both studies, accelerometer and physiological data were collected from the wearables alongside energy expenditure using indirect calorimetry. Three regression algorithms were used to predict metabolic equivalents (METs; ie, random forest, gradient boosting, and neural networks), and five classification algorithms (ie, k-nearest neighbor, support vector machine, random forest, gradient boosting, and neural networks) were used for physical activity intensity classification as sedentary, light, or moderate to vigorous. Algorithms were evaluated using leave-one-subject-out cross-validations and out-of-sample validations.

Results: The root mean square error (RMSE) was lowest for gradient boosting applied to SenseWear and Polar H7 data (0.91 METs), and in the classification task, gradient boost applied to SenseWear and Polar H7 was the most accurate (85.5%). Fitbit models achieved an RMSE of 1.36 METs and 78.2% accuracy for classification. Errors tended to increase in out-of-sample validations with the SenseWear neural network achieving RMSE values of 1.22 METs in the regression tasks and the SenseWear gradient boost and random forest achieving an accuracy of 80% in classification tasks.

Conclusions: Algorithms trained on combined data sets demonstrated high predictive accuracy, with a tendency for superior performance of random forests and gradient boosting for most but not all wearable devices. Predictions were poorer in the between-study validations, which creates uncertainty regarding the generalizability of the tested algorithms.

(JMIR Mhealth Uhealth 2021;9(8):e23938) doi:10.2196/23938

KEYWORDS

bioenergetics; energy balance; accelerometers; machine learning; validation

Introduction

Background

Participation in physical activity results in increased energy expenditure [1] and represents a key modifiable risk factor for cardiovascular disease, obesity, diabetes mellitus, cancer, and mortality [2]. Thus, longitudinal, unobtrusive, and accurate measurement of intraday physical activity energy expenditure would be highly valuable for health research. Activity trackers offer a scalable means for the continuous collection of physical activity data in free-living environments and, by extension, the measurement of energy expenditure. Unfortunately, the accuracy of activity trackers varies greatly between devices and activities [3,4], which limits their use when quantifying energy balance and activity behaviors.

The potential of machine learning techniques to model the complex interactions of accelerometer data, physiological variables, and the rate of energy expenditure has been recognized for some time. Rothney et al [5] trained an artificial neural network using raw accelerometer data as input to predict the energy expenditure in a whole-body calorimetry chamber. Pober et al [6] used quadratic discriminant analysis and a hidden Markov model to classify activity and subsequently estimated the proportion of time performing different activities. Research groups have built on these early findings and have reported highly accurate algorithms for a variety of activities [7-11]. Researchers often take two broad approaches when modeling physical activities: first, attempting to predict the rate of energy expenditure, and second, classifying a minute as sedentary activity, light physical activity, or moderate-to-vigorous physical activity (MVPA), both of which are important for health research. Regression approaches can be used to derive the total energy expenditure for a subject and this can subsequently be incorporated into energy balance models to calculate energy intake [12]. Alternatively, accurately determining the time an individual spends in broader categories of activity or the intensity of that activity can be important for public health guidance. For example, successful weight maintenance in the National Weight Control Registry and weight management recommendations are often defined based on the time an individual spends in MVPA [13]. Machine learning algorithms have the potential to enhance physical activity assessment beyond that of traditional count-based methods, which despite being more accessible, may not be sufficiently accurate for the assessment of energy expenditure and intensity classifications [14].

Recently, we demonstrated in a laboratory validation study that accelerometer and physiological sensor outputs can be modeled using random forests to predict the rate of energy expenditure (as a multiple of resting energy expenditure) in commercial and research-grade activity monitors. We demonstrated a low error in the prediction of energy expenditure [15]. The number of activities in which energy expenditure was measured in this study was limited, and the generalizability of these algorithms remains uncertain. A method for continued refinement of predictive algorithms is to obtain more than one data set [16] to provide larger, more diverse training data with more activities. More data present a new optimization problem, which (because of different assumptions made by different algorithms) means that there is no guarantee that any algorithm will minimize error on all problems [17]. For machine learning models to be used in general health research settings, it is critical to evaluate the generalizability of prediction algorithms. The extent to which an algorithm will generalize is influenced by the characteristics of the sample, activity types, size, and quality of the training data. One approach that addresses each of these limitations is to evaluate prediction algorithms on different samples using data collected under different conditions. In addition to generalizability, a combination of heterogeneous data sets collected under different experimental conditions may help to increase the accuracy of predictions [18].

Objectives

In this study, two distinct data sets of concurrent inputs from multiple wearable devices (ie, Fitbit Charge 2, ActiGraph GT3-x, SenseWear Armband Mini, and a polar chest strap) and measured energy expenditure (indirect calorimetry) are combined to develop predictive models of minute-level energy expenditure and physical activity. We aim to evaluate classification and regression algorithms to (1) predict the rate of energy expenditure and (2) classify a single minute as sedentary activity, light physical activity, or MVPA. Algorithms were validated using leave-one-subject-out cross-validation (LOSO) and out-of-sample validation. Concurrently, we evaluated the SenseWear armband, a device that has been shown to outperform accelerometer-based monitors when classifying activity minutes [19] and is one of the most accurate wrist or arm-based monitors for estimating energy expenditure [3].

Methods

Studies

This study aggregated the data collected as part of two separate studies at the Human Appetite Research Unit, University of Leeds. Participants were recruited from the local area using word-of-mouth and recruitment emails. Participants must have been at least 18 years of age, have been able to attend the research laboratory at the required intervals, be able to ambulate without assistance, they must not have been taking medications known to alter metabolic rate, and participants must not have had any cardiovascular, metabolic, renal disorders, illness, or injury that would increase the risk of medical events during physical activity. Both studies were approved by the University of Leeds, School of Psychology Ethics Committee (PSC-407 and PSC-744 for study 1 and 2, respectively), and all participants provided informed consent before participation in the study. The participant information for the samples is shown in Table 1. Study 2 had proportionately more males, lower age, lower average percentage of fat mass (FM), and a higher resting metabolic rate (RMR) on average.

Table 1. Characteristics of the included sample.

Study	Participants		Age (years), mean (SD)	Height (cm), mean (SD)	Weight[a] (kg), mean (SD)	FFM[b] (kg), mean (SD)	FM[c] (kg), mean (SD)	FM (%), mean (SD)	RMR[d] (kcal/d), mean (SD)
	Total	Female, n (%)							
1	59	41 (69)	44.4 (14.1)	167.5 (8.9)	75.7 (13.6)	49.8 (8.9)	24.8 (10.7)	32.5 (10.3)	1581.8 (280.4)
2[a]	30	13 (43)	31.9 (10.2)	171.9 (9.2)	70.6 (12.9)	55 (12.6)	15.1 (7.1)	21.7 (8.7)	1769.3 (435.8)

[a]In study 2, resting metabolic rate and body composition were estimated at a subsequent visit to the laboratory and therefore weight is not the sum of fat mass and fat-free mass; in study 1, body composition was not available for all subjects and therefore weight is not the sum of fat mass and fat-free mass.

[b]FFM: fat-free mass.

[c]FM: fat mass.

[d]RMR: resting metabolic rate.

Protocols

Study 1

The details of study 1 have been published previously [15]. The protocol of study 1 consisted of 10 activities, each performed for 5 minutes in the following order: sitting, standing, treadmill walking and incline walking (4 km/h), jogging, and incline jogging (6-8 km/h). Participants then rested for 3 minutes and transitioned to a cycle ergometer for low- and moderate-intensity cycling. After another period of recovery, participants performed a folding and sweeping task. Owing to a variation in physical fitness, the jogging task (n=49), incline jogging (n=30), and moderate cycling tasks (n=58) were not performed by all participants.

Study 2

In study 2 (total energy expenditure from wearable devices study), participants visited the lab and refrained from eating or consuming caffeine for at least 4 hours. This exercise visit is the first of three visits to the laboratory conducted as part of a wider project. Weight and height were obtained from a SECA 704s stadiometer and electronic scale (SECA, Germany), and subsequently, an activity protocol was performed. All activities were performed in 5-minute increments, and the order was identical for all participants. First, resting tasks were performed where participants lay supine, sat in a backed chair, and then stood. Next, after a 2-minute unstructured transitional period, participants performed seated typing, standing ironing, and wiping surfaces while standing. After another 2-minute transition, participants walked on a treadmill at 4 km/h, walked at an incline of 5% at 4 km/h, and subsequently jogged at 7 km/h. The participants then rested for 10 minutes. After the unstructured resting period, participants performed low-intensity and moderate-intensity cycling, low-intensity and moderate-intensity rowing, and low-intensity and moderate-intensity cross-training (elliptical), with 1-minute transitions between each, and the intensity of the tasks was determined by a self-selected perceived exertion. In study 2, one participant did not perform rowing or elliptical tasks.

Body Composition Assessment

In both studies, body composition was estimated using air displacement plethysmography (BodPod, Life Measurement, Inc), n=57 in study 1 and n=30 in study 2. Study 2 is part of a wider study in which participants visited the laboratory three times, the first of which was the laboratory validation reported here. Body composition was measured at a subsequent visit to the laboratory in a fasting state.

Energy Expenditure

This study used metabolic equivalents (METs) as the outcome variable, which served to eliminate the proportion of energy expenditure attributable to RMR. We first established the RMR of each participant, which was measured in the fasting state, before any exercise. In both studies, RMR was determined from VO_2 and VCO_2 data collected through a ventilated hood indirect calorimeter system (gas exchange measurement; Nutren Technology Ltd). In study 1, RMR was measured before exercise testing, and in study 2, which occurred on a subsequent visit to the laboratory. After researchers explained the procedures to the participants and an initial calibration process (approximately 10 minutes), VO_2 and VCO_2 were measured for 30 minutes in the supine position. The RMR was established from the VO_2 and VCO_2 of the 5-minute block with the lowest coefficient of variation [20]. If RMR data were unavailable (n=3 across both studies), we approximated the RMR with BMI-specific equations [21]. During the activity sessions, energy expenditure was obtained from a stationary metabolic cart (Vyntus CPX, Jaeger-CareFusion), and these data were expressed relative to the measured RMR of each subject to derive METs. Definitions of METs are inconsistent [22] and we took an individualized approach to METs calculations because the *standard* definition of METs may have limited applicability in some subjects [23].

Devices

Accelerometer and physiological data were collected using various sensors in both protocols. The Polar H7 chest strap (Polar Electro) was used to measure the heart rate. An ActiGraph GT3-X accelerometer (ActiGraph) and a Fitbit Charge 2 (Fitbit Inc) were attached securely to the nondominant wrist. Participants also wore the SenseWear Armband Mini (BodyMedia Inc) on the upper arm.

Data Aggregation

The sensor outputs were obtained from the device-specific software and aggregated to the minute level and time matched to the criterion energy expenditure data. Data loss attributable

to device malfunction was as follows: in study 1, Fitbit data of 2 participants, ActiGraph data of 1 participant, and polar heart rate data of 1 participant were lost. In study 2, 1 SenseWear and 1 Fitbit data set were lost because of device failure. Given the slightly different data availability in each model, our results report the number of minutes used and the number of participants. All minutes in which energy expenditure data were available (ie, face mask was not removed) were included in this analysis, and the aggregation of the data sets by time was conducted in Python 3.7.6 and R version 3.6.3 (R Core Team).

For activity-specific analyses, we grouped activities into broader categories. *Activities of daily living*, which involved folding, sweeping, typing, ironing, and wiping surfaces. Distinct categories were assigned for *cycling*, *elliptical*, *rowing*, *running*, and *walking*. The sedentary activities involved all sitting, standing, and supine tasks. The transitional category refers to unstructured resting or transitional minutes.

Features

Predictive models were built for Fitbit, ActiGraph, and SenseWear, and the features used in each model are listed in Table 2. Each device used a combination of subject-level features, accelerometer features, and physiological features, which have been related to the rate of energy expenditure in previous studies [3,5,24-26]. The features varied depending on the feature availability of each device. Where small (limit of 5 minutes) heart rate gaps existed (eg, loss of signal between the respective heart rate sensor and the skin), we used linear interpolation to fill gaps. As activity in the preceding minutes influences the rate of energy expenditure at the measurement point [27], some time-lagged features were computed: for steps (Fitbit and SenseWear), vector magnitude (ActiGraph), Fitbit heart rate (Fitbit), and polar heart rate (SenseWear and ActiGraph), the change from t-1 minutes for each minute up to t-5 minutes were included as predictive features. In addition, the mean and SD of the current and last 5 minutes were used as predictive features. If time-lagged variables could not be computed due to missing data (ie, for the first minutes for each subject), we imputed backward using the next available observation.

As a constant variance is important for some of the algorithms tested in this study, all numeric features were standardized before training using the following formula:

$$z = (x - \mu) / sd \quad \textbf{(1)}$$

where μ and *sd* refer to the variable mean and SD, respectively.

Table 2. Predictive features used in each of the models.

Device[a] and category	Features
Fitbit	
Subject features	Gender, age, height, weight, and sitting heart rate
Acceleration features	Steps features:
	steps mean, steps difference (t-1, t-2, t-3, t-4, and t-5 minutes); steps mean and SD of last 5 minutes
Physiological features	Fitbit heart rate features:
	Fitbit heart rate above sitting heart rate, Fitbit heart rate percentage of maximum heart rate, Fitbit heart rate mean, Fitbit heart rate difference (t-1, t-2, t-3, t-4, and t-5 minutes), and Fitbit heart rate mean and SD of last 5 minutes
ActiGraph	
Subject features	Gender, age, height, and weight
Acceleration features	X, Y, Z features:
	minimum, maximum, mean, SD; median crossings; 10th, 25th, 50th, 75th, 90th percentiles; correlations (XY, XZ, YZ); dominant frequency; dominant frequency magnitude
	First order differential of X, Y, Z features:
	minimum, maximum, mean, SD; median crossings; 10th, 25th, 50th, 75th, and 90th percentiles; correlations (XY, XZ, YZ); dominant frequency; dominant frequency magnitude
	Vector magnitude features:
	vector magnitude mean; vector magnitude difference (t-1, t-2, t-3, t-4, and t-5 minutes); vector magnitude mean and SD of last 5 minutes
Physiological features	Polar heart rate features:
	polar heart rate above sitting heart rate; polar heart rate percentage of maximum heart rate; polar heart rate mean; polar heart rate difference (t-1, t-2, t-3, t-4, and t-5 minutes); polar heart rate mean and SD of last 5 minutes
SenseWear	
Subject features	Gender, age, height, and weight
Acceleration features	X, Y, Z features:
	peaks, mean of absolute differences, average;
	Steps features:
	steps mean; steps difference (t-1, t-2, t-3, t-4, and t-5 minutes); steps mean and SD of last 5 minutes
Physiological features	Polar heart rate features:
	polar heart rate above sitting heart rate; polar heart rate percentage of maximum heart rate; polar heart rate mean; polar heart rate difference (t-1, t-2, t-3, t-4, and t-5 minutes); polar heart rate mean and SD of last 5 minutes; and SenseWear sensors: near body temperature average, Galvanic skin response average, skin temperature average

[a]For each device, the subject characteristics, acceleration features, and physiological features are listed.

Algorithms

The SenseWear outputs a MET estimate that we evaluated in this study (SenseWear manufacturer). We also tested several machine learning algorithms for regression and classification tasks, which are described below. In the regression tasks, algorithms predicted a MET value for each minute, and in the classification tasks, algorithms classified activity categories for each minute. The activity classifications were as follows: sedentary activity (≤1.5 METs), light physical activity (>1.5 and <3 METs), and MVPA (≥3.0 METs) [18,28,29]. For each algorithm, the hyperparameters were informed by a random search through a range of potential hyperparameters in the preliminary tuning experiments. Random search iterates over a grid of randomly selected combinations of hyperparameters, rather than exploring every possible combination of features, and therefore offers a significant computational advantage over a grid-search approach [30]. Each random search was conducted with the RandomizedSearchCV class in Scikit Learn [31], using three-fold cross-validation. The specific parameters for each algorithm are detailed in Multimedia Appendix 1, and except for the neural network models (explained in the following section), the scoring or loss criterion was the default loss or scoring metrics within Scikit Learn. All algorithms were trained using Keras-GPU [32] or Scikit Learn [31].

Random Forest

The random forest algorithm was used for regression and classification tasks [33]. Random forests involve training of multiple decision trees on data subsamples. Importantly, when splitting these decision trees, only a subsample of the potential predictors is used, which serves to *decorrelate* the trees. The predictions of each tree can then be combined to produce a majority vote (classification) or continuous prediction (regression). The optimal hyperparameters of the algorithm were estimated in the tuning experiments and included the number of trees, number of samples required to split a tree, number of samples per leaf, total predictors, and the depth of

trees. In regression, the quality of a split was assessed with mean square error, and in classification, Gini impurity was used. Algorithms were implemented using the *RandomForestClassifier* and *RandomForestRegressor* classes in Scikit Learn [31].

Gradient Boosting

For the regression and classification tasks, we used the gradient boosting algorithm. Similar to random forests, this algorithm is a tree-based ensemble method. However, where random forests may be considered to use a *bagging* approach, gradient boosting uses *boosting* to learn. Boosting involves the sequential growth of small (weak) decision trees. Each tree is trained using the residuals of the previous estimator and subsequently added to the fitted function to update the residuals. In the boosting phase, a learning rate parameter penalizes the contribution of each tree to the overall model, thereby slowing the learning [34]. The gradient boosting hyperparameters were tuned in the random search experiments and included the number of boosting stages, the maximum depth of the estimators, learning rate, number of samples required to split a node, the number of samples per leaf, and the maximum number of predictors. In the regression, the loss function was least squares, and in classification, deviance was used. Algorithms were implemented using the *GradientBoostingClassifier* and *GradientBoostingRegressor* classes in Scikit Learn [31].

Neural Networks

The third algorithm, used in both regression and classification tasks, was artificial neural networks. Neural networks allow complex, nonlinear functions to be modeled and comprise layers of interconnected *neurons*. At each neuron, inputs are subjected to a numerical activation function, and then passed through subsequent hidden layers of neurons to an output layer [34,35]. In the training process, the interneuronal weights of the network are refined relative to a loss function (ie, mean square error or cross-entropy). Neural networks in the classification studies used the sparse categorical cross-entropy loss function, and in the regression setting, the loss was the mean square error. We tuned the learning rate of each network, the number of layers, and the number of neurons. Neural networks hidden layers used the *relu* activation function, and classification models used a *softmax* activation in the output layer, both classification and regression networks used the Adam optimizer.

K-Nearest Neighbors

For classification tasks, we tested the k-nearest neighbor (KNN) algorithm. This algorithm assigns a given point to a particular class based on the majority class of the k nearest neighbors, where the neighbors of a given point are defined by a distance metric (ie, Euclidian, Minkowski, or Manhattan) [34]. Hyperparameters adjusted in the training process included the number of neighbors in each neighborhood (k), distance metrics, and the weight applied to each of the observations in a neighborhood. KNN was implemented with Scikit Learn [31], using the *KNeighborsClassifier* class.

Support Vector Machine

The final classification model tested was a support vector machine classifier with a radial basis function [35]. A support vector machine aims to find a separating hyperplane between classes by maximizing the distance between the points and the hyperplane. In this study, we tuned the regularization parameter (C) and gamma, which defines the magnitude of the effect of specific training examples. The support vector machine classifier was implemented with the *SVC* class in Scikit Learn [31].

Statistical Analyses

We conducted two validation approaches for all the analyses and algorithms. First, LOSO validations, where algorithms are trained on all but the data of 1 participant, and the participant is held back for validation. This process was repeated until all participants had served as the validation participant once. Second, we used an out-of-sample validation in which the entire data set from one study was used as training data, and the second study was used as an out-of-sample validation. Regression algorithms were evaluated by root mean square error (RMSE), mean absolute percentage error (MAPE) with the *Metrics* package in R and concordance correlation coefficient (CCC) with *DescTools*. Agreement statistics were calculated at the minute level; however, for visualization purposes, we computed the RMSE at the level of individuals and plotted these values. Equivalence tests were used to determine if the true METs and predicted METs were statistically equivalent; tests used equivalence bounds of 10%, and to be considered equivalent, the 90% CI must fall within the equivalence bounds. Finally, linear mixed models with a random intercept of subject ID were used to investigate differences in RMSE between the models. Comparisons were conducted using the Lme4 [36] package in R, with *P* values adjusted by the Bonferroni method in post hoc comparisons. For classification tasks, we report the κ statistic, which compares the accuracy of the predictions to that of a random system. We also report accuracy, where accuracy is the proportion of cases that were classified correctly and the F1 score. All classification statistics were calculated using the Caret [37] package in R. A *P* value of <.05 was used to determine statistical significance, where *P* values were reported.

Results

Regression

A total of 89 participant activity sessions were included in this sample, and all models could be evaluated on at least 5448 minutes of data in the LOSO validations.

The regression algorithms predicting energy expenditure are presented for minute-level data in Table 3 and are visually displayed in Figure 1. Our results demonstrate that the greatest error in METs was observed for the manufacturer-provided SenseWear estimates, with MAPE and RMSE values of 34.54 and 1.86, respectively. For ActiGraph, the RMSE was lowest for gradient boosting (0.93 METs), which also achieved the lowest MAPE of any ActiGraph model (17.88%). Of the Fitbit models, the random forest and gradient boosting had equal RMSE (1.36 METs), but a slightly lower MAPE was achieved by the random forest. For the SenseWear, the gradient boost had the lowest RMSE value (0.91 METs), and this was the lowest RMSE of all those tested. The neural network models were associated with a greater overall RMSE for the ActiGraph, Fitbit, and SenseWear models.

Activity-specific MET predictions are presented in Multimedia Appendix 2, and the RMSE is shown in Figure 2. For all activities tested, tree-based models (gradient boost or random forest) applied to ActiGraph or SenseWear data were superior, as measured by RMSE. The manufacturer estimates of SenseWear had the highest RMSE for all activities aside from sedentary activities, in which only the ActiGraph gradient boost and random forest had a lower RMSE. Notably, all Fitbit models overestimated sedentary activities and had the highest RMSE in this category. The pairwise comparisons between models are presented in Multimedia Appendix 3 for each of the comparisons shown in Figure 1 and Figure 2. An example of the model predictions for a single subject is shown in Figure 3.

Table 4 shows the statistics for the between-study predictions. Notably larger errors were observed relative to the LOSO validations, with the Fitbit gradient boost reaching a RMSE of 1.92 METs (neural network) when study 1 was used as the training data. To estimate the relative importance of each of the features used in each model, permutation importance has been reported in Multimedia Appendix 4.

Table 3. Leave-one-subject-out cross-validation results for each of the regression models.

Model	Minutes[a]	Participants, n (%)	Predicted (METs[b]), mean (SD)	True (METs), mean (SD)	MAPE[c]	RMSE[d]	CCC[e] (95% CI)	Equivalence
SWA[f] manufacturer	5533	88 (99)	3.8 (2.49)	4.04 (2.59)	34.54	1.86	0.73 (0.72-0.74)	__[g]
AG[h] gradient boost	5517	87 (98)	4.04 (2.35)	4.04 (2.59)	17.88	0.93	0.93 (0.93-0.93)	Equivalent[i]
AG neural network	5517	87 (98)	4.05 (2.55)	4.04 (2.59)	21.65	1.14	0.9 (0.9-0.91)	Equivalent
AG random forest	5517	87 (98)	4.05 (2.32)	4.04 (2.59)	18.36	0.94	0.93 (0.92-0.93)	Equivalent
FB[j] gradient boost	5448	86 (97)	4.03 (2.19)	4.01 (2.58)	30.22	1.36	0.84 (0.83-0.84)	Equivalent
FB neural network	5448	86 (97)	4.02 (2.28)	4.01 (2.58)	32.27	1.45	0.82 (0.82-0.83)	Equivalent
FB random forest	5448	86 (97)	4.03 (2.14)	4.01 (2.58)	30.10	1.36	0.84 (0.83-0.84)	Equivalent
SWA gradient boost	5492	87 (98)	4.04 (2.39)	4.04 (2.6)	17.83	0.91	0.93 (0.93-0.94)	Equivalent
SWA neural network	5492	87 (98)	4.05 (2.47)	4.04 (2.6)	19.56	0.96	0.93 (0.92-0.93)	Equivalent
SWA random forest	5492	87 (98)	4.05 (2.35)	4.04 (2.6)	18.25	0.92	0.93 (0.93-0.93)	Equivalent

[a]Minutes refers to the number of minutes the algorithms are validated on.

[b]METs: metabolic equivalents.

[c]MAPE: mean absolute percentage error.

[d]RMSE: root mean square error.

[e]CCC: concordance correlation coefficient CCC is presented with 95% CIs.

[f]SWA: SenseWear.

[g]The model is not statistically equivalent to the criterion.

[h]AG: ActiGraph.

[i]Equivalent implies that the model is statistically equivalent to the criterion.

[j]FB: Fitbit.

Figure 1. Boxplots demonstrating the root mean square error overall for each of the tested models. AG: ActiGraph; FB: Fitbit; RMSE: root mean square error; SWA: SenseWear.

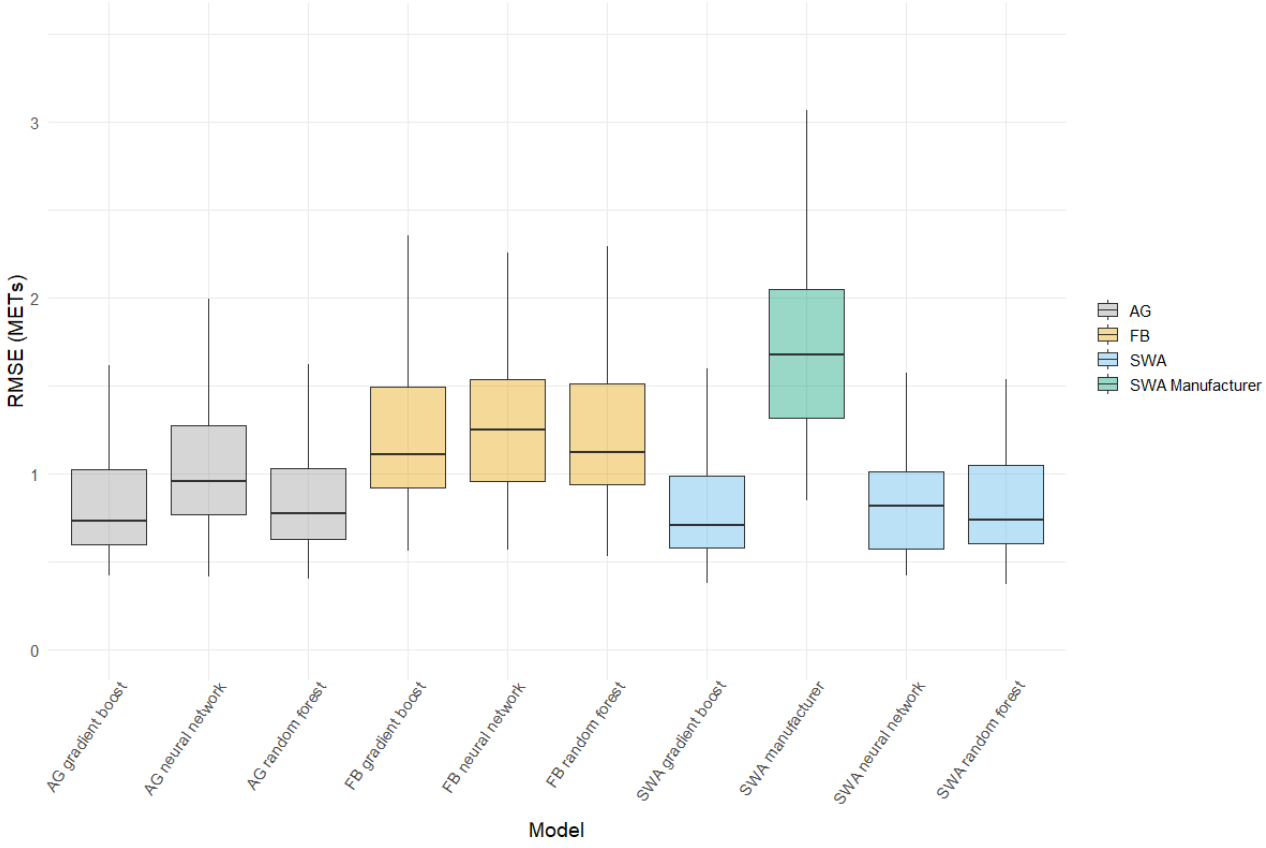


Figure 2. Boxplots demonstrating the root mean square error for each of the tested models in specific activity categories. ADL: activities of daily living; AG: ActiGraph; FB: Fitbit; RMSE: root mean square error; SWA: SenseWear.

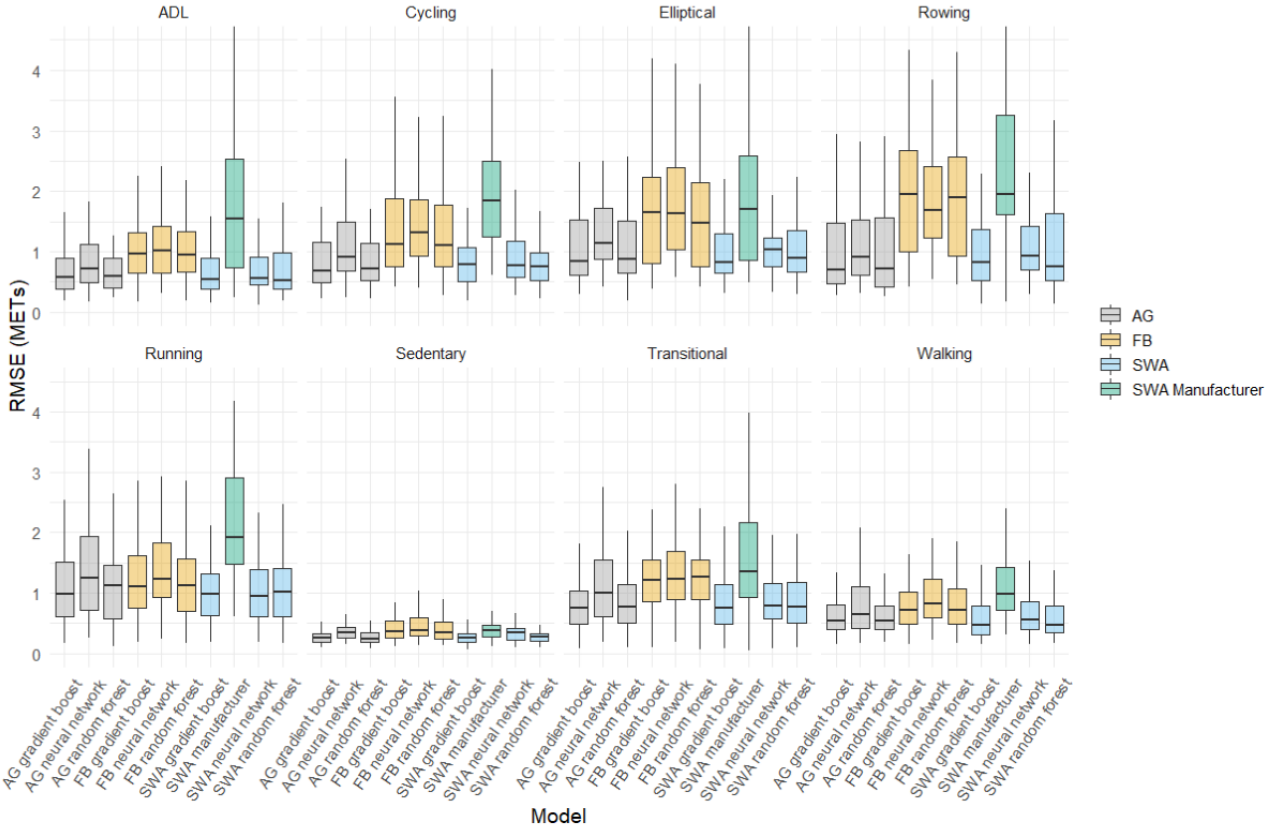


Figure 3. A time series plot showing metabolic equivalents predicted by the models tested in this study (colored solid line) and indirect calorimeter (black dashed line), for a single subject in study 2. The x-axis represents the measurement time. Minutes 1-15=sedentary; minutes 16-17=transitional/unstructured; minutes 18-32=activities of daily living (typing, wiping surfaces, and ironing); minutes 33-34=transitional/unstructured; minutes 35-44=walking; minutes 45-49=running; minutes 50-59=transitional/unstructured; minutes 60-69=cycling; minutes 71-80=rowing; and minutes 82-91=elliptical. Participants performed cycling, rowing, and elliptical tasks at self-selected low and moderate intensity for 5 minutes each. AG: ActiGraph; FB: Fitbit; METs: metabolic equivalents; SWA: SenseWear.

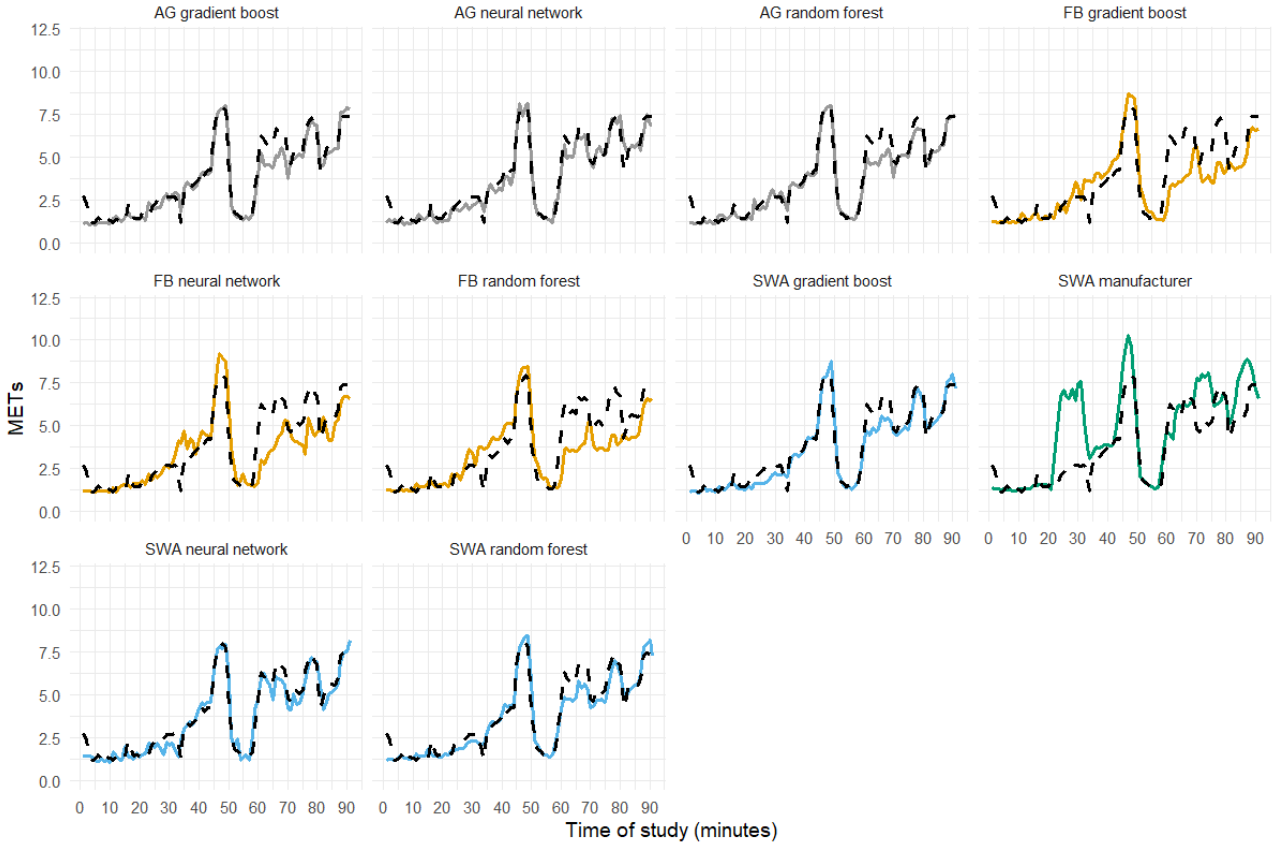


Table 4. Out-of-sample results for each of the regression models.

Model	Training data	Minutes[a]	Predicted (METs[b]), mean (SD)	True (METs), mean (SD)	MAPE[c]	RMSE[d]	CCC[e] (95% CI)	Equivalence
AG[f] gradient boost	Study 1	2690	4.03 (1.9)	3.93 (2.66)	36.35	1.37	0.82 (0.81-0.83)	Equivalent[g]
AG neural network	Study 1	2690	4.07 (2.48)	3.93 (2.66)	29.75	1.33	0.87 (0.86-0.88)	Equivalent
AG random forest	Study 1	2690	3.97 (1.79)	3.93 (2.66)	39.50	1.51	0.78 (0.77-0.79)	Equivalent
FB[h] gradient boost	Study 1	2630	3.76 (1.7)	3.88 (2.65)	47.55	1.89	0.64 (0.62-0.66)	Equivalent
FB neural network	Study 1	2630	3.65 (1.86)	3.88 (2.65)	47.40	1.92	0.65 (0.63-0.67)	—[i]
FB random forest	Study 1	2630	3.76 (1.66)	3.88 (2.65)	47.45	1.87	0.64 (0.63-0.66)	Equivalent
SWA[j] gradient boost	Study 1	2633	3.92 (2.13)	3.94 (2.68)	27.35	1.23	0.87 (0.86-0.88)	Equivalent
SWA neural network	Study 1	2633	3.88 (2.26)	3.94 (2.68)	27.07	1.22	0.88 (0.87-0.89)	Equivalent
SWA random forest	Study 1	2633	3.91 (2.07)	3.94 (2.68)	29.54	1.28	0.86 (0.85-0.87)	Equivalent
AG gradient boost	Study 2	2827	4.46 (2.14)	4.15 (2.52)	31.49	1.36	0.83 (0.82-0.84)	—
AG neural network	Study 2	2827	4.24 (2.56)	4.15 (2.52)	29.00	1.42	0.84 (0.83-0.85)	Equivalent
AG random forest	Study 2	2827	4.45 (2.1)	4.15 (2.52)	31.47	1.38	0.82 (0.81-0.84)	—
FB gradient boost	Study 2	2818	4.11 (2.06)	4.13 (2.51)	34.38	1.66	0.74 (0.72-0.75)	Equivalent
FB neural network	Study 2	2818	4.01 (2.04)	4.13 (2.51)	33.10	1.56	0.77 (0.75-0.78)	Equivalent
FB random forest	Study 2	2818	4.21 (2.04)	4.13 (2.51)	33.79	1.62	0.75 (0.73-0.77)	Equivalent
SWA gradient boost	Study 2	2859	4.15 (2.13)	4.14 (2.51)	24.90	1.25	0.86 (0.85-0.87)	Equivalent
SWA neural network	Study 2	2859	3.94 (2.36)	4.14 (2.51)	25.65	1.25	0.87 (0.86-0.88)	Equivalent
SWA random forest	Study 2	2859	4.2 (2.13)	4.14 (2.51)	25.72	1.26	0.85 (0.84-0.86)	Equivalent

[a]Minutes refers to the number of minutes the algorithms are validated on.

[b]METs: metabolic equivalents.

[c]MAPE: mean absolute percentage error.

[d]RMSE: root mean square error.

[e]CCC: concordance correlation coefficient CCC is presented with 95% CIs.

[f]AG: ActiGraph.

[g]Equivalent implies that the model is statistically equivalent to the criterion.

[h]FB: Fitbit.

[i]The model is not statistically equivalent to the criterion.

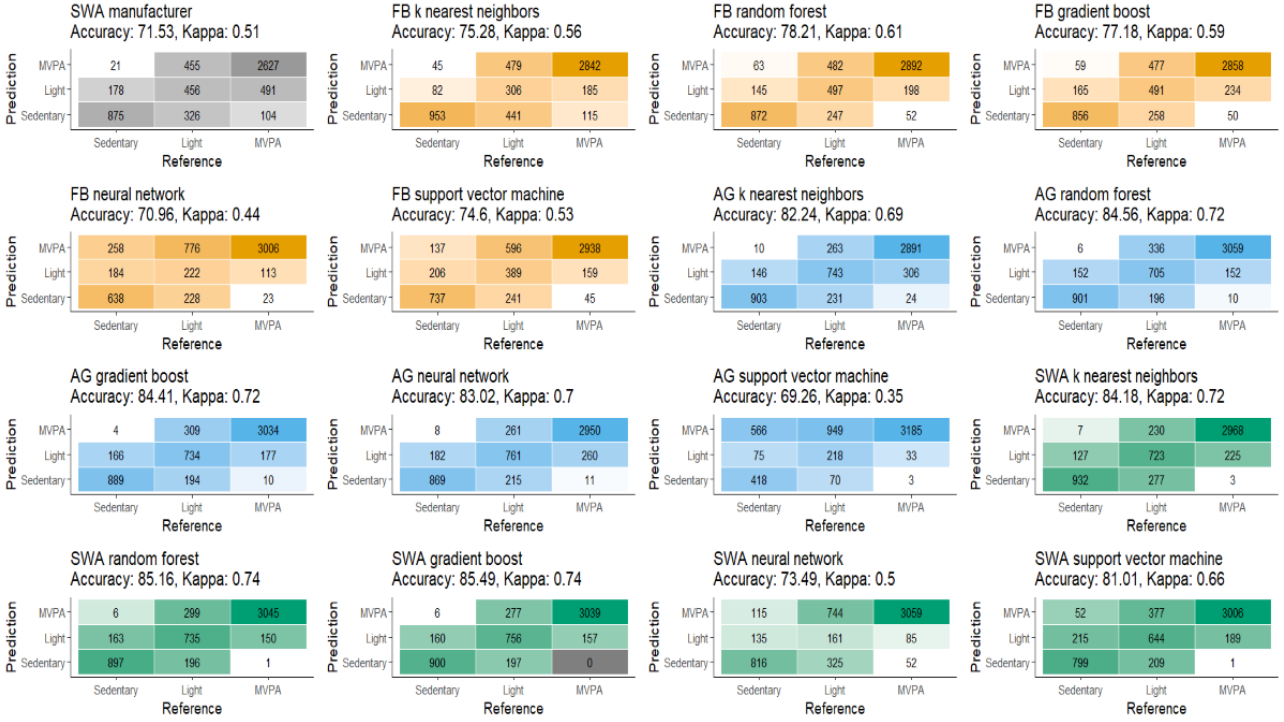
[j]SWA: SenseWear.

Classification

Figure 4 presents the results of the LOSO classification experiments for all classification algorithms and the SenseWear manufacturer estimates. Classes were slightly imbalanced, approximately 19.4% sedentary activity, 22.4% light physical activity, and 58.2% MVPA with small differences between devices due to data availability. The highest accuracy for Fitbit models was the random forest (78.21%), for the ActiGraph models, the random forest achieved the highest accuracy (84.56%), and for the SenseWear models, the gradient boosting algorithm (85.49%) was the most accurate.

Multimedia Appendix 5 provides class-specific statistics for each model. Models tended to perform worse in light activity with F1 scores ranging from 0.20 (SenseWear neural network) to 0.66 (SenseWear gradient boost). In sedentary activities, the F1 score was improved with a range of 0.54 (Actigraph support vector machine) to 0.83 (four models). For MVPA, the F1 score ranged from 0.80 (Actigraph support vector machine) to 0.93 (three models).

Figure 4. A confusion matrix detailing the classification accuracies for each of the tested models. AG: ActiGraph; FB: Fitbit; SWA: SenseWear.



Between-Study Predictions

The between-study classification accuracies are listed in Table 5. In most cases, when study 1 served as the training data, lower accuracy was observed. When study 1 served as the training data, the accuracy ranged from 0.55 (ActiGraph support vector machine) to 0.80 (two models). When study 2 served as the training data, the accuracy ranged from 0.65 (ActiGraph support vector machine) to 0.79 (three models).

Table 5. Between-study classification results for each of the classification models.

Training data and model	Accuracy	κ
Study 1		
AG[a] gradient boost	0.75	0.55
AG k-nearest neighbors	0.61	0.35
AG neural network	0.72	0.52
AG random forest	0.74	0.53
AG support vector machine	0.55	0.06
FB[b] gradient boost	0.67	0.43
FB k-nearest neighbors	0.68	0.47
FB neural network	0.67	0.47
FB random forest	0.67	0.41
FB support vector machine	0.67	0.45
SWA[c] gradient boost	0.80	0.67
SWA k-nearest neighbors	0.74	0.57
SWA neural network	0.79	0.66
SWA random forest	0.80	0.66
SWA support vector machine	0.68	0.43
Study 2		
AG gradient boost	0.79	0.56
AG k-nearest neighbors	0.72	0.48
AG neural network	0.75	0.51
AG random forest	0.79	0.57
AG support vector machine	0.65	0.07
FB gradient boost	0.73	0.48
FB k-nearest neighbors	0.72	0.47
FB neural network	0.71	0.44
FB random forest	0.73	0.48
FB support vector machine	0.73	0.48
SWA gradient boost	0.78	0.57
SWA k-nearest neighbors	0.76	0.55
SWA neural network	0.76	0.55
SWA random forest	0.79	0.58
SWA support vector machine	0.78	0.55

[a]AG: ActiGraph.

[b]FB: Fitbit.

[c]SWA: SenseWear.

Discussion

Principal Findings

This study aggregated two laboratory data sets to build on previous work demonstrating the potential for machine learning algorithms to produce accurate estimates of METs and intensity classes in a diverse set of activities and participants. In both regression and classification settings, we observed the smallest errors in energy expenditure predictions when applying tree-based algorithms (ie, random forest and gradient boosting) to SenseWear and ActiGraph outputs with the RMSE and classification errors generally being higher for Fitbit models. In almost all cases, the error was smaller than the SenseWear manufacturer estimates, and in out-of-sample generalizability experiments, we observed greater error and lower accuracy when compared with the LOSO validations. We believe that

this is the first study to classify the intensity of activity using machine learning algorithms in Fitbit devices. In Fitbit models, we demonstrated accuracies up to approximately 78% (κ=0.6), with superior performance observed for sedentary activity and MVPA classifications, but these were generally less accurate than ActiGraph and SenseWear models, where up to approximately 85% accuracy (κ=0.74) was achieved. Taken together, and if these results are verified in free-living, ecologically valid examples, these findings imply that highly accurate estimates of energy expenditure, sedentary activity, and MVPA behaviors can be estimated by the wearables tested here.

Algorithm Accuracy

We used neural networks, random forests, and gradient boosting in regression tasks. In previous studies, neural networks and random forests have been shown to be effective in modeling energy expenditure [8,9], and our results confirm this to an extent. The RMSE values observed in the trained models ranged from 0.91 METs to 1.45 METs, which improve upon the SenseWear manufacturer value of approximately 1.86 METs. However, when the average METs in this study were considered (approximately 4 METs), it was evident that the energy expenditure prediction could be further improved. It is noteworthy that neural networks resulted in the highest RMSE for all 3 devices and performed particularly poorly for Fitbit models. Similarly, Kate et al [38] showed that neural networks resulted in bias significantly different from 0, compared with bagged decision trees and numerous other algorithms, which were not statistically different. Despite the utility of deep neural networks to model highly nonlinear functions, in some use cases, the *no free lunch* theorems broadly state that there will not be an optimal algorithm for all tasks [17]. Indeed, for our data sets, tree-based ensemble models are generally superior for both learning tasks. It may be that a higher RMSE can be reduced by larger training sets [39].

We generated lagged accelerometer and heart rate variables for each model because the rate of energy expenditure depends on the rate of work in preceding minutes [27], and the relative importance of these metrics is evidenced in the variable importance analyses. Including time-lagged features allows for a clearer distinction between minutes that are relatively similar in their accelerometer pattern but differ in their measured energy expenditure, that is, sitting for a prolonged period versus sitting immediately after running. Transitional minutes were on average approximately 3 METs (largely attributable to the activity in the preceding minutes), compared with sedentary minutes, which averaged approximately 1.3 METs, yet the error statistics were generally comparable with those observed in sedentary minutes, indicating that algorithms could distinguish between those minutes. More advanced neural network architectures (ie, recurrent neural networks) [40] may further the ability of models to capture the temporal dependencies of energy expenditure.

Generalization

Although many studies have reported low errors when using machine learning approaches in the estimation of energy expenditure or classification of activity, external (out-of-sample) validations are rarer and the opportunity to identify cases of overfitting has been limited. Therefore, we used an out-of-sample validation between the two data sets. In all cases, we observed performance degradation when compared with the LOSO validations. Some of this reduction in accuracy is probably attributable to differences in protocols, activities, and participants, which means that algorithms do not have *similar* minutes on which to train. In addition, it is possible that the algorithms overfit the data. Overfitting occurs when a complex model learns the *noise* in the training data, which does not represent the true underlying function between the inputs and the output [41]. Previous studies have used out-of-sample validation or validation in free-living environments [10,42,43], and when compared with laboratory validations, errors may increase. Concerning the classification of physical activity intensity in multiple samples, a previous study reported reductions in out-of-sample accuracy relative to the within-sample validated models, in some algorithm and data set comparisons [44]. However, the machine learning models still outperformed the Euclidean norm minus one GGIR classification method in out-of-sample testing. In another comprehensive generalizability study, five lab-based heterogeneous data sets were used to predict exercise intensity. This study found that when models were applied to a different data set than those they were generated on, model accuracy decreased from 72-95% to 41-60% [18]. These drops are notably higher than those in this study, and this is probably attributable to the greater differences in the accelerometer models, wear position, and samples across the five data sets. However, caution must be exercised in a comparison between studies, as the balance of classes is likely to differ and therefore influence some evaluation metrics.

Classification

Our LOSO validations demonstrated a relatively high predictive accuracy (75-85%). However, research-grade device models (ActiGraph and SenseWear) were superior. Fitbit devices provide estimates of time in each category (ie, sedentary, light, and MVPA), but the criteria and algorithms remain proprietary. Feehan et al [45] compared estimates of time in intensities with devices such as ActiGraph and Actical, and concluded that more than 80% of studies reported errors >10% with mean differences ranging between 44% and 632% for estimations of activity above light intensity. Importantly, the devices used for comparison in many studies have varying cut points and are not necessarily gold standards. Our results indicate that the application of machine learning to intensity classification can refine the large errors observed in previous studies. Despite the promising results, we emphasize that laboratory studies have limited ecological validity, and future research should seek to address this. Whole-room indirect calorimetry would likely allow more realistic behaviors to be studied while providing a gold standard comparator.

Strengths and Limitations

A strength of this study is the aggregation of two data sets to provide a more comprehensive and variable data set on which to train models, although the measures (sensors and indirect calorimetry) were the same between studies. The tested cohorts differed demographically, and the protocols were heterogeneous,

which provides a good estimate of the applicability of the tested models. Combining data sets also leads to a larger number of participants (n=89), which is a larger sample size than much of the previous literature [7,9,10,44,46,47]. In general, an increase in training observations is considered a mechanism for enhancing performance [41], and the results of this study provide some evidence that this is the case in both commercial and research-grade accelerometers.

Another strength of this study is the testing of numerous algorithm and device combinations. A previous study developed a multilayer neural network that was trained on a wearable system including a vest for electrocardiogram measurements and 4 accelerometers (one on each wrist and thigh) [47]. Despite the small bias, this is unlikely to be a feasible means of assessing free-living energy balance behaviors. Participant discomfort and sensor removal present additional biases (ie, missing data), which may require additional modeling approaches to address [48-50]. The threshold of practicality varies depending on the size, duration, computational resources, and specific aims of the research study. Therefore, the development of three models with varying requirements is a central advantage of this study.

Testing both classification and regression algorithms in the same devices enhances the use of the results of this study. One area of future work is to explore combined classification and regression approaches, similar to the branched models of the Actiheart [51] or stacked ensemble approaches. This may be effective in producing refined estimates of total daily energy expenditure in free-living subjects, given that most of a day comprises resting or sedentary minutes and some of our models slightly overestimate sedentary activities, although depending on the classification or regression methods, this could incur additional computational costs when applying this to larger data sets. Future work in our lab will examine the application of such models to free-living environments against a doubly labeled water criterion.

A limitation of this study is the lack of a true testing set. Rather, we attempt to develop an unbiased estimate of the true test error by (1) testing on unseen participants and (2) testing on an unseen data set. In the former, the within-subject data are generally more correlated than the between-subject data, and this method represents the closest approximation of how such a model would perform in practice [8]. In the latter, this is extended so that the training and testing sets comprised different participants and protocols. Beyond these validation approaches, the ultimate test of the results presented here is a free-living validation for energy expenditure and intensity classes. The total daily energy expenditure can be validated using the doubly labeled water method over a 7- to 14-day period [52], and the results presented in this paper are part of a wider project in which we aim to validate model predictions in free-living. Although free-living validations are critical, the resolution required to evaluate activity-specific errors can only be obtained from indirect calorimetry. Regarding activity categories, no gold standard method exists to validate time in sedentary activity, light physical activity, and MVPA outside of a controlled environment, and the generalizability of classification models to free-living studies is somewhat uncertain. The authors have highlighted the limitations of accelerometer data collected within a laboratory [53,54]; the activities performed in a free-living environment are more diverse, which further necessitates the need for more naturalistic (ie, free-living) validation studies or at least validation studies conducted over several days using diverse activity protocols in a residential facility. Next, to replicate predictions made by the present algorithms in free-living subjects, measured RMR may be required, which increases the researcher and participant burden. A suitable alternative in the absence of measured RMR would be prediction equations derived from BMI, age, height, and gender, rather than assuming a resting value of 3.5 ml O_2/kg/min [55,56]. Finally, our use of the measured RMR to calculate *METs* may contribute to differences between the tested algorithms and the SenseWear manufacturer.

Conclusions

This study builds on previous work from our lab and others, demonstrating that machine learning techniques can be used to learn the complexities of human movement and physiological data in the study of human energy expenditure. Classification and regression errors were greater when comparisons were made between studies. Single-sample, cross-sectional studies generating energy expenditure models show acceptable accuracy; however, it is likely that these models are overfitted to a given sample, and thus, improving generalizability is essential. To extend the utility of energy expenditure estimates beyond lab conditions, more cross testing between data sets is required, in addition to validation in free-living samples by doubly labeled water.

Acknowledgments

This research was funded by an Engineering and Physical Sciences Research Council impact acceleration grant (EP/R511717/1).

Authors' Contributions

ROD, JT, MH, GF, CD, and RJS designed the study. ROD and JT collected the data. ROD, JT, and GWH analyzed the data. ROD, JT, MH, CD, GWH, GF, and RJS contributed to writing and reviewing the manuscript.

Conflicts of Interest

RJS consults for Slimming world UK through Consulting Leeds, which is a wholly owned subsidiary of the University of Leeds. The other authors declare no conflicts of interest.

Multimedia Appendix 1
Hyperparameters used in each of the models.
[DOCX File , 30 KB - mhealth_v9i8e23938_app1.docx]

Multimedia Appendix 2
Leave-one-subject-out cross-validation results for each of the regression models in each of the activity categories.
[DOCX File , 32 KB - mhealth_v9i8e23938_app2.docx]

Multimedia Appendix 3
Between-model comparisons for root mean square error in each of the tested activity types.
[DOCX File , 69 KB - mhealth_v9i8e23938_app3.docx]

Multimedia Appendix 4
Permutation importance analysis for Fitbit, SenseWear, and Actigraph datasets.
[DOCX File , 124 KB - mhealth_v9i8e23938_app4.docx]

Multimedia Appendix 5
Leave-one-subject-out cross-validation results for each of the classification models in each of the intensity categories.
[DOCX File , 22 KB - mhealth_v9i8e23938_app5.docx]

References

1. Westerterp KR. Physical activity and physical activity induced energy expenditure in humans: measurement, determinants, and effects. Front Physiol 2013;4:90 [FREE Full text] [doi: 10.3389/fphys.2013.00090] [Medline: 23637685]
2. Warburton DE, Nicol CW, Bredin SS. Health benefits of physical activity: the evidence. Can Med Assoc J 2006 Mar 14;174(6):801-809 [FREE Full text] [doi: 10.1503/cmaj.051351] [Medline: 16534088]
3. O'Driscoll R, Turicchi J, Beaulieu K, Scott S, Matu J, Deighton K, et al. How well do activity monitors estimate energy expenditure? A systematic review and meta-analysis of the validity of current technologies. Br J Sports Med 2020 Mar;54(6):332-340. [doi: 10.1136/bjsports-2018-099643] [Medline: 30194221]
4. Shcherbina A, Mattsson C, Waggott D, Salisbury H, Christle JW, Hastie T, et al. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. J Pers Med 2017 May 24;7(2):3 [FREE Full text] [doi: 10.3390/jpm7020003] [Medline: 28538708]
5. Rothney MP, Neumann M, Béziat A, Chen KY. An artificial neural network model of energy expenditure using nonintegrated acceleration signals. J Appl Physiol (1985) 2007 Oct;103(4):1419-1427 [FREE Full text] [doi: 10.1152/japplphysiol.00429.2007] [Medline: 17641221]
6. Pober DM, Staudenmayer J, Raphael C, Freedson PS. Development of novel techniques to classify physical activity mode using accelerometers. Med Sci Sports Exerc 2006 Sep;38(9):1626-1634. [doi: 10.1249/01.mss.0000227542.43669.45] [Medline: 16960524]
7. Staudenmayer J, Pober D, Crouter S, Bassett D, Freedson P. An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. J Appl Physiol (1985) 2009 Oct;107(4):1300-1307 [FREE Full text] [doi: 10.1152/japplphysiol.00465.2009] [Medline: 19644028]
8. Ellis K, Kerr J, Godbole S, Lanckriet G, Wing D, Marshall S. A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers. Physiol Meas 2014 Nov;35(11):2191-2203 [FREE Full text] [doi: 10.1088/0967-3334/35/11/2191] [Medline: 25340969]
9. Montoye AH, Begum M, Henning Z, Pfeiffer KA. Comparison of linear and non-linear models for predicting energy expenditure from raw accelerometer data. Physiol Meas 2017 Feb;38(2):343-357. [doi: 10.1088/1361-6579/38/2/343] [Medline: 28107205]
10. Ellis K, Kerr J, Godbole S, Staudenmayer J, Lanckriet G. Hip and wrist accelerometer algorithms for free-living behavior classification. Med Sci Sports Exerc 2016 May;48(5):933-940 [FREE Full text] [doi: 10.1249/MSS.0000000000000840] [Medline: 26673126]
11. Ahmadi MN, Chowdhury A, Pavey T, Trost SG. Laboratory-based and free-living algorithms for energy expenditure estimation in preschool children: a free-living evaluation. PLoS One 2020;15(5):e0233229 [FREE Full text] [doi: 10.1371/journal.pone.0233229] [Medline: 32433717]
12. Shook RP, Hand GA, O'Connor DP, Thomas DM, Hurley TG, Hébert JR, et al. Energy intake derived from an energy balance equation, validated activity monitors, and dual X-Ray absorptiometry can provide acceptable caloric intake data among young adults. J Nutr 2018 Mar 1;148(3):490-496. [doi: 10.1093/jn/nxx029] [Medline: 29546294]
13. Ostendorf DM, Lyden K, Pan Z, Wyatt HR, Hill JO, Melanson EL, et al. Objectively measured physical activity and sedentary behavior in successful weight loss maintainers. Obesity 2018 Jan;26(1):53-60 [FREE Full text] [doi: 10.1002/oby.22052] [Medline: 29090513]

14. Lyden K, Kozey SL, Staudenmeyer JW, Freedson PS. A comprehensive evaluation of commonly used accelerometer energy expenditure and MET prediction equations. Eur J Appl Physiol 2011 Feb;111(2):187-201 [FREE Full text] [doi: 10.1007/s00421-010-1639-8] [Medline: 20842375]
15. O'Driscoll R, Turicchi J, Hopkins M, Horgan GW, Finlayson G, Stubbs JR. Improving energy expenditure estimates from wearable devices: a machine learning approach. J Sports Sci 2020 Jul;38(13):1496-1505. [doi: 10.1080/02640414.2020.1746088] [Medline: 32252598]
16. Chowdhury AK, Tjondronegoro D, Chandran V, Trost SG. Ensemble methods for classification of physical activities from wrist accelerometry. Med Sci Sports Exerc 2017 Sep;49(9):1965-1973. [doi: 10.1249/MSS.0000000000001291] [Medline: 28419025]
17. Wolpert DH, Macready WG. No free lunch theorems for optimization. IEEE Trans Evol Computat 1997 Apr;1(1):67-82. [doi: 10.1109/4235.585893]
18. Farrahi V, Niemela M, Tjurin P, Kangas M, Korpelainen R, Jamsa T. Evaluating and enhancing the generalization performance of machine learning models for physical activity intensity prediction from raw acceleration data. IEEE J Biomed Health Inform 2020 Jan;24(1):27-38. [doi: 10.1109/jbhi.2019.2917565]
19. Calabró MA, Lee J, Saint-Maurice PF, Yoo H, Welk GJ. Validity of physical activity monitors for assessing lower intensity activity in adults. Int J Behav Nutr Phys Act 2014 Sep 28;11:119 [FREE Full text] [doi: 10.1186/s12966-014-0119-7] [Medline: 25260625]
20. Sanchez-Delgado G, Alcantara JM, Ortiz-Alvarez L, Xu H, Martinez-Tellez B, Labayen I, et al. Reliability of resting metabolic rate measurements in young adults: Impact of methods for data analysis. Clin Nutr 2018 Oct;37(5):1618-1624. [doi: 10.1016/j.clnu.2017.07.026] [Medline: 28826698]
21. Müller MJ, Bosy-Westphal A, Klaus S, Kreymann G, Lührmann PM, Neuhäuser-Berthold M, et al. World health organization equations have shortcomings for predicting resting energy expenditure in persons from a modern, affluent population: generation of a new reference standard from a retrospective analysis of a German database of resting energy expenditure. Am J Clin Nutr 2004 Nov;80(5):1379-1390. [doi: 10.1093/ajcn/80.5.1379] [Medline: 15531690]
22. Byrne NM, Hills AP, Hunter GR, Weinsier RL, Schutz Y. Metabolic equivalent: one size does not fit all. J Appl Physiol 2005 Sep;99(3):1112-1119 [FREE Full text] [doi: 10.1152/japplphysiol.00023.2004] [Medline: 15831804]
23. Melzer K, Heydenreich J, Schutz Y, Renaud A, Kayser B, Mäder U. Metabolic equivalent in adolescents, active adults and pregnant women. Nutrients 2016 Jul 20;8(7) [FREE Full text] [doi: 10.3390/nu8070438] [Medline: 27447667]
24. Brage S, Brage N, Franks PW, Ekelund U, Wareham NJ. Reliability and validity of the combined heart rate and movement sensor Actiheart. Eur J Clin Nutr 2005 Apr;59(4):561-570. [doi: 10.1038/sj.ejcn.1602118] [Medline: 15714212]
25. Whybrow S, Ritz P, Horgan G, Stubbs R. An evaluation of the IDEEA™ activity monitor for estimating energy expenditure. Br J Nutr 2012 Apr 2;109(1):173-183. [doi: 10.1017/s0007114512000645]
26. Ceesay SM, Prentice AM, Day KC, Murgatroyd PR, Goldberg GR, Scott W, et al. The use of heart rate monitoring in the estimation of energy expenditure: a validation study using indirect whole-body calorimetry. Br J Nutr 1989 Mar;61(2):175-186. [doi: 10.1079/bjn19890107] [Medline: 2706223]
27. McArdle WD, Katch FI, Katch VL. Exercise Physiology: Nutrition, Energy, and Human Performance. Baltimore, USA: Lippincott Williams & Wilkins; 2010.
28. Blair CK, Morey MC, Desmond RA, Cohen HJ, Sloane R, Snyder DC, et al. Light-intensity activity attenuates functional decline in older cancer survivors. Med Sci Sports Exerc 2014 Jul;46(7):1375-1383 [FREE Full text] [doi: 10.1249/MSS.0000000000000241] [Medline: 24389524]
29. Beaulieu K, Hopkins M, Blundell J, Finlayson G. Impact of physical activity level and dietary fat content on passive overconsumption of energy in non-obese adults. Int J Behav Nutr Phys Act 2017 Feb 6;14(1):14 [FREE Full text] [doi: 10.1186/s12966-017-0473-3] [Medline: 28166797]
30. Ge´ron A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems 2nd Edition. Sebastopol, CA: O'Reilly Media; 2019.
31. Pedregosa F, Weiss R, Brucher M, Varoquaux G, Gramfort A, Michel V. Scikit-learn: machine learning in python. J Mach Learn Res 2011:2825-2830 [FREE Full text]
32. Chollet F. Keras-team/keras: Deep Learning for Humans. GitHub. 2015. URL: https://github.com/keras-team/ [accessed 2020-06-05]
33. Breiman L. Random Forests. Mach Learn 2001;9:5-32 [FREE Full text] [doi: 10.1023/A:1010933404324]
34. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. New York, NY: Springer Series in Statistics; 2009.
35. Kuhn M, Johnson K. Applied Predictive Modeling. New York, NY: Springer; 2013.
36. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using. J Stat Soft 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
37. Kuhn M. Building predictive models in R using the caret package. J Stat Soft 2008;28(5):1-26. [doi: 10.18637/jss.v028.i05]
38. Kate RJ, Swartz AM, Welch WA, Strath SJ. Comparative evaluation of features and techniques for identifying activity type and estimating energy cost from accelerometer data. Physiol Meas 2016 Mar;37(3):360-379 [FREE Full text] [doi: 10.1088/0967-3334/37/3/360] [Medline: 26862679]

39. DeGregory KW, Kuiper P, DeSilvio T, Pleuss JD, Miller R, Roginski JW, et al. A review of machine learning in obesity. Obes Rev 2018 May;19(5):668-685 [FREE Full text] [doi: 10.1111/obr.12667] [Medline: 29426065]
40. Paraschiakos S, de Sá SC, Okai J, Slagboom EP, Beekman M, Knobbe A. RNNs on Monitoring Physical Activity Energy Expenditure in Older People. arXiv. 2020. URL: https://tinyurl.com/cfp7849a [accessed 2021-07-06]
41. Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. PLoS One 2019;14(11):e0224365 [FREE Full text] [doi: 10.1371/journal.pone.0224365] [Medline: 31697686]
42. Willetts M, Hollowell S, Aslett L, Holmes C, Doherty A. Statistical machine learning of sleep and physical activity phenotypes from sensor data in 96,220 UK Biobank participants. Sci Rep 2018 May 21;8(1):7961 [FREE Full text] [doi: 10.1038/s41598-018-26174-1] [Medline: 29784928]
43. Sasaki JE, Hickey AM, Staudenmayer JW, John D, Kent JA, Freedson PS. Performance of activity classification algorithms in free-living older adults. Med Sci Sports Exerc 2016 May;48(5):941-950 [FREE Full text] [doi: 10.1249/MSS.0000000000000844] [Medline: 26673129]
44. Montoye AH, Westgate BS, Fonley MR, Pfeiffer KA. Cross-validation and out-of-sample testing of physical activity intensity predictions with a wrist-worn accelerometer. J Appl Physiol 2018 May 1;124(5):1284-1293 [FREE Full text] [doi: 10.1152/japplphysiol.00760.2017] [Medline: 29369742]
45. 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 9;6(8):e10527 [FREE Full text] [doi: 10.2196/10527] [Medline: 30093371]
46. Zhang S, Rowlands AV, Murray P, Hurst TL. Physical activity classification using the GENEA wrist-worn accelerometer. Med Sci Sports Exerc 2012 Apr;44(4):742-748. [doi: 10.1249/MSS.0b013e31823bf95c] [Medline: 21988935]
47. Lu K, Yang L, Seoane F, Abtahi F, Forsman M, Lindecrantz K. Fusion of heart rate, respiration and motion measurements from a wearable sensor system to enhance energy expenditure estimation. Sensors 2018 Sep 14;18(9) [FREE Full text] [doi: 10.3390/s18093092] [Medline: 30223429]
48. Lee P. Data imputation for accelerometer-measured physical activity: the combined approach. Am J Clin Nutr 2013 May;97(5):965-971. [doi: 10.3945/ajcn.112.052738] [Medline: 23553165]
49. Xu SY, Nelson S, Kerr J, Godbole S, Patterson R, Merchant G, et al. Statistical approaches to account for missing values in accelerometer data: applications to modeling physical activity. Stat Methods Med Res 2018 Apr;27(4):1168-1186. [doi: 10.1177/0962280216657119] [Medline: 27405327]
50. O'Driscoll R, Turicchi J, Duarte C, Michalowska J, Larsen SC, Palmeira AL, et al. A novel scaling methodology to reduce the biases associated with missing data from commercial activity monitors. PLoS One 2020;15(6):e0235144 [FREE Full text] [doi: 10.1371/journal.pone.0235144] [Medline: 32579613]
51. Brage S, Brage N, Franks PW, Ekelund U, Wong M, Andersen LB, et al. Branched equation modeling of simultaneous accelerometry and heart rate monitoring improves estimate of directly measured physical activity energy expenditure. J Appl Physiol 2004 Jan;96(1):343-351 [FREE Full text] [doi: 10.1152/japplphysiol.00703.2003] [Medline: 12972441]
52. Black AE, Cole TJ. Within- and between-subject variation in energy expenditure measured by the doubly-labelled water technique: implications for validating reported dietary energy intake. Eur J Clin Nutr 2000 May;54(5):386-394. [doi: 10.1038/sj.ejcn.1600970] [Medline: 10822285]
53. Bastian T, Maire A, Dugas J, Ataya A, Villars C, Gris F, et al. Automatic identification of physical activity types and sedentary behaviors from triaxial accelerometer: laboratory-based calibrations are not enough. J Appl Physiol 2015 Mar 15;118(6):716-722 [FREE Full text] [doi: 10.1152/japplphysiol.01189.2013] [Medline: 25593289]
54. 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]
55. Kim D, Lee J, Park HK, Jang DP, Song S, Cho BH, et al. Comparing the standards of one metabolic equivalent of task in accurately estimating physical activity energy expenditure based on acceleration. J Sports Sci 2017 Jul;35(13):1279-1286. [doi: 10.1080/02640414.2016.1221520] [Medline: 27556835]
56. Kozey SL, Lyden K, Howe CA, Staudenmayer JW, Freedson PS. Accelerometer output and MET values of common physical activities. Med Sci Sports Exerc 2010 Sep;42(9):1776-1784 [FREE Full text] [doi: 10.1249/MSS.0b013e3181d479f2] [Medline: 20142781]

Abbreviations

KNN: k-nearest neighbor
LOSO: leave-one-subject-out cross-validation
MAPE: mean absolute percentage error
MET: metabolic equivalent
MVPA: moderate-to-vigorous physical activity
RMR: resting metabolic rate
RMSE: root mean square error

Edited by L Buis; submitted 28.08.20; peer-reviewed by KF Chen, Z Ren, MS Salekin; comments to author 04.10.20; revised version received 18.11.20; accepted 18.05.21; published 04.08.21.

Please cite as:
O'Driscoll R, Turicchi J, Hopkins M, Duarte C, Horgan GW, Finlayson G, Stubbs RJ
Comparison of the Validity and Generalizability of Machine Learning Algorithms for the Prediction of Energy Expenditure: Validation Study
JMIR Mhealth Uhealth 2021;9(8):e23938
URL: https://mhealth.jmir.org/2021/8/e23938
doi:10.2196/23938
PMID:34346890

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org